Thermodynamic Models from Fluctuation Solution Theory Analysis of Molecular Simulations

Christensen, Steen

Publication date:
2008

Thermodynamic Models from Fluctuation Solution Theory Analysis of Molecular Simulations

Steen Christensen

August 29, 2007

Department of Chemical Engineering
Technical University of Denmark

Preface

This thesis is submitted as partial fulfillment of the requirements for the Ph.D.-degree at Danmarks Tekniske Universitet (Technical University of Denmark). The work has been carried out at Institut for Kemiteknik (Department of Chemical Engineering) from August 2004 to August 2007 under the supervision of Associate Professor Jens Abildskov. I am grateful for the funding for this work, which is partially financed by Danisco A/S, the Graduate School in Chemical Engineering (MP_2T) and DTU. I gratefully acknowledge access to the Danish Center for Scientific Computing at the University of Southern Denmark, Odense, Denmark.

I would like to thank students and coworkers at the Computer Aided Process Engineering Center (CAPEC) at the Department of Chemical Engineering, Technical University of Denmark for making the center a fun and inspiring place.

Special thanks go to Dr. Abildskov for a good and inspiring collaboration where the door always was open when I needed a sparring partner. This work is also a result of collaboration between several departments and I would like to thank Dr. Flemming Y. Hansen and Dr. Günther H. Peters at the Department of Chemistry, University of Denmark for our fruitful discussions through the years. The last piece in the puzzle is Professor John P. O'Connell whom I visited in the spring of 2006 at the University of Virginia and I would like to thank for the hospitality and great inspiration.

Finally I would like to thank my friends and family for helping me through rough periods. I would also like to thank my friends in the cycling community both in Denmark and Virginia for helping me tanking my mind of the project when needed.

Østerbro, Denmark, 2007
Steen Christensen

Abstract

This thesis describes the development of a methodology for generation of G^E-model descriptions for binary mixtures. The methodology addresses the problem of missing descriptions of phase behavior of mixtures containing specialty compounds where both experimental data and G^E-model parameters typically does not exist. The G^E-model parameters are necessary for development of new products. The G^E-model descriptions are generated by combining molecular dynamics simulations and Fluctuation Solution Theory where the only input needed is the pure component force field descriptions. The methodology contains two main steps. In the first step a series of molecular dynamic simulations are made at different compositions from which the average molar volume and angle-averaged center-of-mass to center-of-mass radial distribution functions are the main output. In the second step the output from simulations are analyzed using Fluctuation Solution Theory which takes the microscopic level information and returns macroscopic derivative properties. The macroscopic derivative properties are then applied as observed quantities in an objective function which is minimized in order to develop macroscopic G^E-model descriptions which can be applied in prediction of phase behavior. The methodology has been tested using a series of case studies. The method is remarkably accurate irrespective of its simplicity. The method only requires open source software and can draw upon the CHARMM parameters known from the literature. Attempts to develop force field parameters for a novel class of compounds was also made,. These were only partially successful.

Resumé på dansk

Denne tese beskriver en fremgangsmåde til bestemmelse af makroskopisk G^E-modeller for binære blandinger. Fremgangsmåden er et forsøg på at tackle problemer ved design af nye kemiske produkter, hvor der sjældent eksisterer eksperimentel data eller model beskrivelser for blandinger. Bestemmelsen af G^E-modellen for en binær blanding opnås ved en kombination af molekylesimulering og *Fluctuation Solution Theory*, hvor det eneste inddata er parameter til kraftfelt beskrivelsen af de rene komponenter. Fremgangsmåden kan opdeles i to hovedtrin. I det første foretages en serie af molekylesimuleringer ved forskellige sammensætninger, hvor molar volumner og radiær fordelingsfunktioner (Radial distribution functions) er de primære resultater. I andet trin bruges de radiær fordelingsfunktioner sammen med *Fluctuation Solution Theory* til at ekstrahere information om de makroskopiske værdier, og denne information bruges til at optimere parametrene i G^E-beskrivelsen, som kan bruges til faseberegninger. Fremgangsmåden er blevet testet igennem en række konkrete eksempler.

Contents

Preface iii

Abstract v

Resumé på dansk vii

1 Introduction 1

2 Molecular Modeling 3
2.1 Gibbs Ensemble Monte Carlo 4
2.1.1 TraPPE-UA Force Field 7
2.1.2 Anisotropic United Atom Force Field 9
2.2 COSMO-RS 11
2.3 Gibbs Free Energy Integration 12
2.4 Summation 16

3 Fluctuation Solution Theory 17
3.1 Practical Fluctuation Solution Theory 18
3.2 MD simulations 18
3.2.1 Force Field 19
3.2.2 MD versus MC 20
3.3 RDF from Simulation and their Integration; 3 Methods 21

4 Generation of G^E-model Parameters 27
4.1 Optimization and Objective Functions 27
4.2 Variance of Parameters 29
4.2.1 General Newton-Raphson method 29
4.2.2 Marquardt Method 30
4.2.3 Modified Marquardt Method 31
4.2.4 Co-variance Matrix and uncertainties 32

5 Prediction of Bubble Point Pressure 33
5.1 Treatment of the Vapor Phase 33
5.2 Uncertainty Calculations 35

6 Optimization of Parameters for Force Field Descriptions 37
6.1 Determination of Intramolecular Potential 37
6.2 Determination of Intermolecular Potential 38
6.2.1 Transferability of parameters 39
6.2.2 Partial Charges Determination 39

6.2.3 Time Consumption 40
6.3 Simple Approach 40
6.3.1 Simulations 41
6.3.2 Optimization Procedure 41

7 Case Studies **43**
7.1 Benzene - Methyl Acetate 43
7.2 Methyl acetate - acetone 49
7.3 Benzene - Ethanol 55
7.4 Methyl acetate - n-pentane 60
7.5 Ethanol - (2H)-Heptafluoropropane 64
7.5.1 Mixture Simulations 66
7.5.2 Vapor Phase Options 67
7.6 Lactones 73
7.6.1 Force Field Optimization 73
7.6.2 Lactones with benzene 78
7.6.3 Comment on Lactone Results 86

8 Conclusion **91**

A Fluctuation Solution Theory **95**
A.1 Grand Canonical Ensemble 95
A.2 Radial Distribution Function 96
A.3 Chemical Potential Relation 98
A.4 Advances by O'Connell 100

B FST Supplements **103**
B.1 Inversion of a Matrix 103
B.2 Switch dependency of μ_i from constant V to P 104
B.3 Derivation of properties general case 105
B.4 Derivation of properties for binary case 107
B.5 From Mole Numbers to Mole Fractions 108

C CHARMM Parameter Tables **109**

D Parameters for Reverse Approach Analysis **115**

References **117**

List of Abbreviations **127**

1

Introduction

In the modern chemical industry the main focus has changed from bulk to high-value specialty chemicals. The property prediction tools available for an R&D group are traditionally group contribution methods and equations of state which were developed with bulk chemistry and oil production challenges in mind. The semi-quantitative property prediction models (SQPP) have been applied to the new challenges but they all have weaknesses and limitations concerning size and complexity of the molecules and mixtures when applied outside a narrow proven range. They are also quite data-demanding to develop and extend. In the area of specialty chemicals experimental data is limited. Typically density data is available at ambient temperatures together with melting and boiling temperatures. Other pure component properties such as enthalpy of vaporization and vapor pressure data can be found in limited cases. The lack of experimental data becomes a major problem when different forms of mixture phase equilibria have to be studied. Normally limited or no data is available for specialty chemicals which makes it impossible to determine parameters for any SQPP model. It has been estimated that the cost of a single VLE data point is $2600 and take 2 days to produce (Gubbins and Quirke, 1996) and the prize has probably not decreased in the last 10 years. An alternative to experimental investigations is molecular modeling or computational chemistry where pseudo-experimental data can be extracted from computations.

Computational simulation techniques are already being used to complement, guide and sometimes replace experimental measurements thereby reducing the time and cost of bringing ideas from the lab to practical application. The increased focus on specialty chemicals and the creation of new products, requires more precise control and understanding of physical and chemical properties. Different computational chemistry techniques will increasingly be used to provide the knowledge at several levels that influence macroscopic behavior.

The current state is highly influenced by the availability of computer hardware capable of handling highly complex computations which has increased rapidly over the last decade to support the development and use of advanced computational software. Computational tools (software packages) available today are greatly improved from those developed just a few years ago with new algorithms, new theoretical methods and ability to explore highly parallel computer systems. This includes both commercial products and software distributed under the GNU license. The overall time required for generating data

from computational methods have been reduced. As already mentioned, computer power and better and faster algorithms have decreased the calculation time. But time spent by the engineer on setting up and analyzing simulations can be reduced by systematization of these procedures and different software is need for this task. In the future a major driving force of the chemical industry is the need to design new materials rationally. Modeling and simulation is a critical technology for achieving these goals because it can provide the detailed technical understanding needed to design useful and profitable materials. In the future, models created by multi-disciplinary teams will be developed more quickly and will be more indicative of reality. Together with industrial investments in moderately parallel supercomputers there is a big potential for molecular modeling techniques and tools for aiding chemical engineers in their development of new products. However, results of molecular simulations need to be competitive with those of experiments in terms of accuracy. Also, the use of molecular simulations is still mostly limited to experts. What is needed is a method which is accurate and easy-to-use. Also when species of contemporary complexity are involved. A method also should be possible to use with cheap and readily software.

In this work a methodology is proposed for generation of excess Gibbs energy models for binary liquid mixtures. The main aspects is a combination of Fluctuation Solution Theory (FST) and molecular dynamics (MD) simulations. The practical steps of the methodology are described in the following chapters. The setup of simulations and the analysis of the results requires only a basic knowledge of computational chemistry. This makes the methodology a potentially important tool for chemical engineers in R&D groups who traditionally do not have expert knowledge of molecular modeling. The cost of implementing the methodology in R&D departments is very low as only open source software is applied.

In Chapter 2 an overview of state-of-the-art molecular modeling techniques is presented. They cover Gibbs Ensemble Monte Carlo (GEMC) simulations, Gibbs free energy integration and, the software COSMO-RS. The fundamentals of FST are given in Chapter 3 together with the practical steps of how to combine FST and MD simulations. In Chapter 4 the last steps of the methodology is introduced where parameters for G^E-models are generated from the results of the FST analysis of the MD simulation. Chapter 5 shows different approaches for applying the results of the methodology in vapor-liquid equilibrium (VLE) calculations. The MD simulations are based on a mathematical description of the interaction between atoms within a molecule and in between molecules, called a force field. In Chapter 6 various approaches for determining the parameters of a force field are described together with normal concepts and assumptions in the force field generation. The theory described in previous chapters are applied in several case studies in Chapter 7 followed by a conclusion.

2

Molecular Modeling

Molecular modeling (MM) is a term which refers to a group of theoretical methods and computational techniques to model the behavior of molecules. The size of the molecular systems studied range from small chemical systems to large biological molecules and material assemblies. Three methods are covered in this chapter of which two have been established within the area of prediction of thermodynamic properties where they have shown top results within the Industrial Fluid Phase Simulation Challenge (IFPSC) program (Case *et al.*, 2004, 2005). Two competing GEMC force fields TraPPE-UA and AUA4 are presented in Section 2.1 after a general introduction to GEMC methods (Panagiotopoulos, 1987; Martin and Siepmann, 1998; Ungerer *et al.*, 2000). The second method is based upon quantum chemical calculations and is now a commercial software package called COSMO*therm* (Klamt, 2005). The third method is Gibbs Free Energy Integration which has shown great potential but also has certain limitations (Chialvo and Haile, 1987).

After an introduction to the method and a review of some of the published results, it is the intention to give an unbiased overview of capabilities, usefulness, and practical utility from the perspective of a chemical engineer with some basic insight into molecular modeling.

It has always been discussed if calculated properties from MM should be a supplement to or substitute of experimental data. Experience has shown that while MM can generally be used for decisions in the initial design phase of a separation process, experimental validation is needed before investments are made in process equipment. The cost of experiments also plays an important role, however among most multinational chemical companies laboratory facilities are available and it does not represent any problems to make the experiments. There are two groups of compounds where MM computations are preferable. The first group is toxic compounds where special facilities are needed together with restrictions and constraints on the use of experimental apparatus which makes the experiments expensive and problematic. The main characteristic of the second group of compounds is their natural occurrence in dilute form which makes the expenditures for purification extremely high. When making systematic experimental studies the starting point is compounds in their purest form. Pure component properties can by determined without loosing or polluting the compound sample. The study of mixture properties however needs larger quan-

tities which makes experiments expensive especially if several mixtures has to be covered in the studies. The second compound group covers flavoring agents and fragrances, but can also cover small active ingredients in pharmaceuticals and agricultural products, or unwanted byproducts which need to be removed.

2.1 Gibbs Ensemble Monte Carlo

The GEMC methodology basically makes it possible to simulate or sample coexisting phases and their properties. When two or more phases are in equilibrium they have the same temperature and pressure and the chemical potential of each component is the same in all phases as shown below for a two-phase system.

$$T^I = T^{II} \tag{2.1}$$
$$P^I = P^{II} \tag{2.2}$$
$$\mu_i^I = \mu_i^{II}, i = 1 \ldots \nu \tag{2.3}$$

The GEMC method makes it possible to achieve the $2 + \nu$ equilibrium criteria. In its original formulation the ensemble was divided into two boxes and considered under NVT-conditions (Panagiotopoulos, 1987). These boxes are considered to be representing bulk phases of two phases in equilibrium as shown in the top of Figure 2.1. The two phase chambers are not in any direct contact but surrounded by mirror images of themselves. As there is no direct contact, surface tension issues and diffusion through phase boundaries do not have to be considered. There are three types of moves considered in the GEMC methodology which are shown in Figure 2.1. The first step is a displacement of a particle within a phase. The second step type is change in volume, ΔV, in volume of one the boxes, which is perfectly correlated with a volume change in the other box as the total volume is constant (Panagiotopoulos *et al.*, 1988). This step type is commonly known as volume rearrangement. The third type of step is interchange of particles, where a randomly selected particle is moved from one phase to the other.

The first type of step is made to ensure equilibrium within each phase or simulation box. The second type of step ensures an equilibrium pressure between the two phases as required by Equation 2.2. The third type of step ensures equilibrium in the free energy or chemical potential for each component between the phases, as shown in Equation 2.3. The equal temperatures are insured by the ensemble where the temperature is kept constant.

This simple macroscopic explanation of the ensemble was used in the original article (Panagiotopoulos, 1987). Soon after it was elaborated to explain the methodology from a statistical point of view (Panagiotopoulos *et al.*, 1988).

The introduction of molecules which could not be simplified to Lennard-Jones particles introduced new parameters and new challenges. In the initial work only united atoms models were considered. United atoms is a simplification where hydrogen atoms are united with the bigger atom to which they are

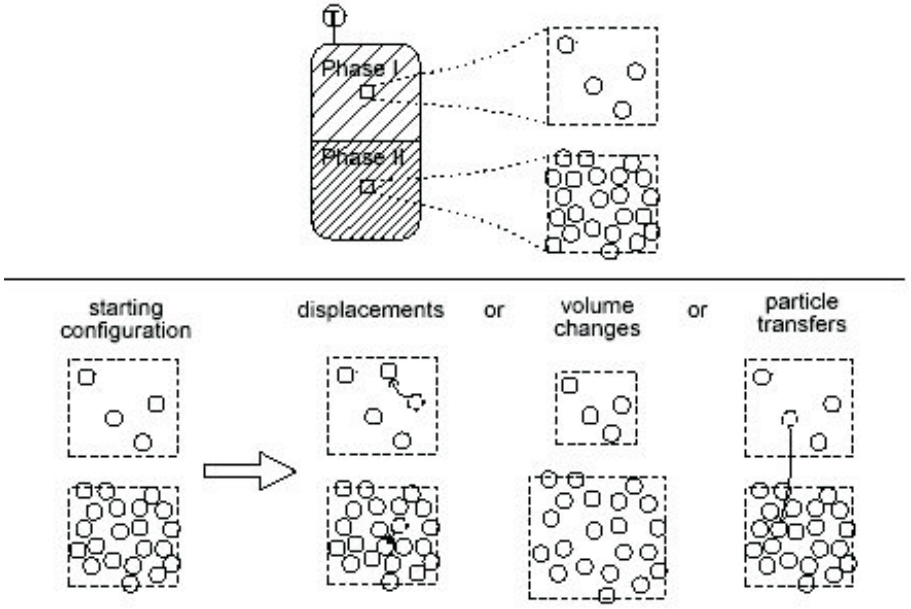


Figure 2.1. Scheme showing the basic assumption of the Gibbs Ensemble Monte Carlo method and the three possible steps (Panagiotopoulos, 1992).

bonded and thereby generating a bead (Laso *et al.*, 1992). This method was first applied to alkanes of various lengths but later it has been expanded to cyclic and branched molecules (Wick *et al.*, 2000; Chen *et al.*, 2001). The insertion or movement of a molecule into a dense phase was in the early work declared practically impossible (Panagiotopoulos *et al.*, 1988). This is because of the probability of randomly finding void space to place the molecule (where the void extends in the randomly selected orientation of the molecule) is very small. But the Configurational-bias Monte Carlo strategy changed this viewpoint where the molecule is inserted in a segment manner (Laso *et al.*, 1992; Vlugt *et al.*, 1992). The method was later modified to handle strongly associating molecules such as acids and alcohols which have a tendency to form clusters because of hydrogen bonding (Chen and Siepmann, 2000) and the latest addition enables simulation of cyclic compounds (Errington and Panagiotopoulos, 1999b; Martin and Siepmann, 1999; Bourasseau *et al.*, 2002a). The impressive list of advances is both a strength and weakness of the method. A lot of issues have been dealt with, which makes it possible to simulate studied compound groups. Yet one can never be sure, the developed insertion moves can handle a compound group considered novel with respect to GEMC. The wide range of possible insertion moves also makes it difficult to setup the simulations where a distribution or probability of selecting each MC move in the GEMC simulations has to be given. One will have to conduct extensive analysis of initial simulations

before the correct distribution of MC moves have been obtained and the actual simulations can be executed.

The simulation of co-existing phases made it possible to predict a range of new properties, where the densities of the coexisting phases together with the vapor pressure curve are the most common. Critical properties are also predicted but statistical noise, due to the sizes of the system normally simulated, makes it infeasible to find critical properties directly. They are determined indirectly using a procedure where the subcritical data is extended using the density scaling law for the critical temperature, the law of rectilinear diameters for the critical density, and Clasius-Clapeyron equation for the critical pressure (Martin and Siepmann, 1999). In the GEMC framework energies of transfer can be calculated for example the heat of vaporization (Wick *et al.*, 2003). The numerous properties where predictive capability have been made available is a major strength of the GEMC method, especially the vapor pressure.

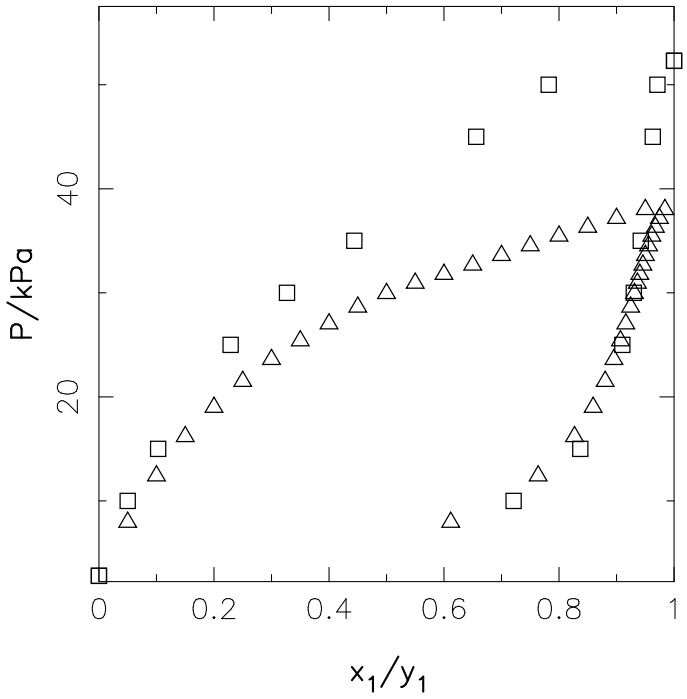


Figure 2.2. Diethyl ether (1) - ethanol (2) at 283.15K comparing phase diagram predicted using GEMC by Stubbs *et al.* (2004) (□) with experimental data of Nagai and Isii (1935) (Δ).

For multicomponent systems it is possible to make an NPT ensemble where the volumes of the two boxes vary independently (Smit and Frenkel, 1989). Stubbs *et al.* (2004) made binary VLE GEMC simulations for the systems diethyl ether - ethanol at 283.15 K (Figure 2.2) and acetone - n-hexane at

328.15K using the TraPPE-UA force field however the results are not particularly accurate. Later Lopes and Tildesley (1997) have made an expansion to multiphase equilibria where several subsystems of simulation boxes are used. A limited number of publications present predictions for binary mixtures of condensable species where one of the components is not a gas (like CO_2) which could indicate two weaknesses (Potoff *et al.*, 1999; Nath and de Pablo, 2000; Wick *et al.*, 2000; Stubbs *et al.*, 2001; Chen *et al.*, 2001; Stubbs *et al.*, 2004; Wick *et al.*, 2005; Kamath *et al.*, 2006). It is difficult, exhaustive and takes a lot of work and simulation time to generate each data point. The results are not good possibly because errors of the pure component vapor pressures propagate to the binary region. An example is shown in Figure 2.2 where the error in the prediction of pure diethyl ether vapor pressure (at $x_1 = 1$) effects most of the phase diagram as the predicted mixture bubble points are too high. One can also question the need to include the vapor phase in the simulation of binary mixtures. At ambient pressures the non-ideality of the vapor phase is insignificant, which makes the large effort needed for GEMC simulation of binary mixtures, unnecessary.

The TraPPE-UA and AUA4 work presented below can be categorized as GEMC force fields because their parameters have been optimized by reproduction of pure component thermodynamic properties made accessible by the GEMC methods. They represent two different approaches of representing the molecules and also two sets of simulation software. From the view of a third party academic or industrial researcher it is challenging that a common platform has not been developed where all major contributors to GEMC make their advances in algorithms and simulation techniques available. A good attempt in making a common plat has been developed by Martin and Siepmann (1999)[1] which will be discussed below.

2.1.1 TraPPE-UA Force Field

Martin and Siepmann (1998, 1999) have developed a force field based on GEMC method named "Transferable Potentials for Phase Equilibria - United Atom" which is normally reduced to TraPPE-UA.

The united atom model is characterized by constant bond lengths and the intramolecular potential is reduced to bond angles and dihedral angles, though there are a special contribution to describe proximity effects between oxygen atoms in glycols (Stubbs *et al.*, 2004). The Lennard-Jones potential is used to describe the van der Waals forces and a simple term to describe interaction due to charges.

$$\begin{aligned} U(\mathbf{R}) &= \sum_{angles} \frac{k_\theta}{2}(\theta - \theta_0) \\ &\quad + \sum_{dihedrals} c_0 + c_1[1 + \cos(\phi)] + c_2[1 - \cos(2\phi)] + c_3[1 + \cos(3\phi)] \end{aligned}$$

[1] http://towhee.sourceforge.net

$$+ \sum_{nonbond} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \left\{ \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\} \tag{2.4}$$

k_θ is the bond angle force constant, θ is the bond angle and θ_0 is its equilibrium position. c_i are the Fourier coefficients of the dihedral or torsional angle potential and ϕ is the dihedral angle. ϵ_{ij} and σ_{ij} are the Lennard-Jones well depth and size, q_i and q_j are the partial charges, ϵ_0 is the effective dielectric constant and r_{ij} is the distance between beads i and j.

The transferable part of the force field name refers to the objective to transfer parameters regressed for specific compound class to partially describe parts of molecules from a different compound class. An example could be 1-pentanol where the alcohol parameters consists of the hydrogen and hydroxyl oxygen plus the alpha carbon group to which it is connected. The parameters for the rest of the beads in the molecule are taken from the alkane description. The procedure for implementing the transferability is by sequentially optimizing the parameters starting with alkanes and moving on to for example alkenes and alcohols in a fashion resembling the way group contribution models like UNIFAC were developed (Fredenslund *et al.*, 1975). Martin and Siepmann (1998) initiated their work with linear alkanes and the database of parameters have been expanded with branched alkanes (Martin and Siepmann, 1999), aromatic compounds (Wick *et al.*, 2000), alcohols both primary, secondary and tertiary (Chen *et al.*, 2001), other oxygen containing compound (Stubbs *et al.*, 2004) and nitrogen containing compounds (Wick *et al.*, 2005). The large number of compound groups covered by the TraPPE force field is impressive and the large effort put into covering many compound groups and atom types makes TraPPE versatile. The large parameter set available makes it possible to get good initial estimates of parameters for a new compound group.

The optimization of the parameters for the TraPPE-UA force field can roughly be divided into two parts, intramolecular and intermolecular parameters. Intramolecular parameters have, if possible, been taken from published force fields or have been determined through *ab initio* calculations (Martin and Siepmann, 1998; Wick *et al.*, 2005). The intermolecular parameters have been optimized by reproduction of liquid densities and critical properties (critical temperature, T_c, critical density, ρ_c, and, critical pressure, P_c) which were estimated by extrapolation of subcritical coexisting data as described above.

A great strength of TraPPE-UA is the speed of calculations caused by the simplified united atom molecular description, however this characteristic has some shortcomings (Chen and Siepmann, 1999). Among several problems the most significant is the prediction of too high vapor pressures which result in too low boiling point temperatures while critical properties are reproduced correctly (Martin and Siepmann, 1998, 1999). Chen and Siepmann (1999) developed an Explicit Hydrogen (TraPPE-EH) version of the force field where hydrogen atoms are included but in a simplified manner reducing the cost of computations and the vapor pressure problem was reduced.

The greatest strength of TraPPE-UA is the easy accessibility through the

open source software package, Towhee (Martin and Thompson, 2004). All the algorithms designed to handle different challenges encountered during the development of the force field and the GEMC methods are included. Together with an extensive manual and example directory the combination of TraPPE and Towhee is the best available option for novices from academia or industry who wants to initiate MM and GEMC studies.

2.1.2 Anisotropic United Atom Force Field

Ungerer *et al.* (2000) was like Siepmann *et al.* interested in the capabilities of the GEMC method and he based his force field on the work of Toxvaerd (1990, 1997). Toxvaerd (1990) was not satisfied with the alkane simulation results of united atom (UA) force fields for pressure and heat of vaporization as function of temperature and density. The all-atom (AA) force fields where all atoms are considered were dismissed as being too expensive in their computational evaluation as the number of van der Waals centers would increase by a factor 3. Instead he proposed the anisotropic united atom (AUA) force field where the van der Waals interaction center is moved from the carbon atom center to the geometric center of the methylene or methyl groups by a vector $\mathbf{R}_i$ which is displayed in Figure 2.3. The position of the interaction center depends on

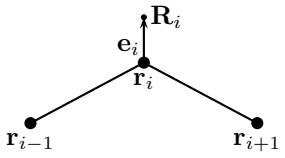


Figure 2.3. Displacement of van der Waals force center in anisotropic force field. $\mathbf{r}_j$ are the carbon atom centers.

two carbon centers for the methyl group, $\mathbf{R}_i(r^2)$ and three for the methylene group, $\mathbf{R}_i(r^3)$. For simplicity the centers of mass of the anisotropic united atom was kept at the center of the carbon atom. The intramolecular forces is reduced to bond angle and dihedral angle the bond lengths are kept constant. The Lennard-Jones potential is used to describe the van der Waals interactions using Lorenz-Berthelot combining rules.

$$\begin{aligned} U(\mathbf{R}) &= \sum_{angles} \frac{1}{2} k_\theta \left(\cos\theta - \cos\theta_0\right)^2 + \sum_{dihedrals} \sum_{j=0}^{8} a_j \left(\cos\Phi\right)^j \\ &+ \sum_{nonbond} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right] \end{aligned} \tag{2.5}$$

k_θ is the bond angle force constant, θ is the bond angle and θ_0 is its equilibrium position. a_i are the dihedral angle potential coefficients and Φ is the dihedral angle. ϵ_{ij} and σ_{ij} are the Lennard-Jones well depth and size, r_{ij} is the distance between interaction centers i and j. The torsional potential differs significantly from other potentials by having a much lower gauche to gauche energy barrier which permits chain rotations at room temperature (Toxvaerd, 1990). The bond angle potential differs also from other work though it is still harmonic. The calculations are accelerated as the inverse cosine does not have to be calculated which is the case in most force fields (see Eq. 2.4 for TraPPE-UA).

The AUA force field by Ungerer *et al.*, named AUA4, has several characteristics which makes it superior to force fields like TraPPE-UA when it comes to computational speed. Even though extra calculations are needed to locate the L-J interaction centers the calculation time of the intermolecular potential is not increased significantly as the number of L-J interaction centers is the same. The mathematical construction of the angle bend and dihedral angle potentials accelerates the calculations as the inverse cosine does not have to be calculated. The AUA4 has the same advantages as TraPPE-UA when compared to AA force fields.

Ungerer *et al.* (2000) formulated an objective function, SS_{AUA4}, for determination of the intermolecular parameters of the AUA4 force field.

$$SS_{AUA4} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(X_i^{mod} - X_i^{exp}\right)^2}{s_i^2} \tag{2.6}$$

where X_i^{exp} is an experimental property which can be liquid density, vapor pressure or enthalpy of vaporization. X_i^{mod} is the computed value of the ith property with the estimated statistical uncertainty s_i^2, ideally variance. The optimization ends when the partial derivative of all m parameters, y_j, are zero.

$$\frac{\partial SS_{AUA4}}{\partial y_j} = \frac{1}{n}\sum_{i=1}^{n}\frac{2\left(X_i^{mod} - X_i^{exp}\right)\frac{\partial X_i^{mod}}{\partial y_j}}{s_i^2} \text{ for } j = 1\ldots m \tag{2.7}$$

The approach is normal in optimization of parameters for mathematical models. The numerical evaluation of derivatives , using forward differences, at each iteration is a formidable computaional effort (Ungerer *et al.*, 2000). The number of simulations at each iteration is roughly equal to $(1 + m) \cdot (n/3)$ which is 1 set of evaluations for the parameter set plus m perturbations where each parameter is modified to get the derivative. Assuming liquid density, vapor pressure and enthalpy of vaporization can be extracted from each simulation $n/3$ simulations have to be performed for each parameter set. Bourasseau *et al.* (2003) came up with method for reducing the simulation time by calculating the derivatives during the normal simulation without increasing the simulation time significantly. The structured and methodical approach to optimizing all parameters of a new compound group is a great advantage which saves the

developer many hours of work. This goes both for analyzing the results and the number of simulations to be performed.

The AUA4 force field shows good reproduction of the thermodynamic properties included in the objective function (Ungerer *et al.*, 2000; Bourasseau *et al.*, 2002a,b). AUA4 shows great advantages with respect to vapor pressure compared to TraPPE-UA without including additional L-J centers which makes it superior. Bourasseau *et al.* (2002a) presented parameters and developments in the GEMC techniques to simulated cyclic alkanes which has not been covered by other groups, although cyclohexane has been simulated by Errington and Panagiotopoulos (1999b) using a similar technique. Bourasseau *et al.* (2002b) developed a crank-shaft MC move based on old work by Verdier and Stockmayer (1962) which makes it possible to model and correctly sample large molecules.

The few simulation results available for binary mixtures covers gas-liquid mixtures where Henry constants have been determined (Ungerer *et al.*, 2000; Boutard *et al.*, 2005). Binary vapor-liquid equilibria studies have not published which makes comparison impossible.

The big drawback of AUA4 and the MC moves and techniques developed together with it, is the inaccessibility of the simulation software. It is very difficult to access with academic work in mind and impossible for any commercial interests because no simulation soft package is not available on the market.

2.2 COSMO-RS

The COSMO framework differs from other contributions to this field of research by being based upon Quantum Mechanical (QM) calculations (Klamt, 1995). While the molecule is considered in vacuum in normal QM calculations the COSMO use continuum solvation models which makes it possible to describe the molecule under liquid-like conditions (Eckert and Klamt, 2002). The calculated molecule surface is divided into segments of area, a_i, with a Screening Charge Density (SCD), σ_i, and a probability distribution of the SCD's, $p^X(\sigma)$, is calculated for compound X. The distributions are called "σ-profiles" and they are the computational demanding part of the COSMO method, but they only have to be calculated once for each compound.

When a mixture of a certain composition has to be treated the solution σ-profile, $p_S(\sigma)$, is found by averaging.

$$p_S(\sigma) = \sum_{i \in S} x_i p^{X_i}(\sigma) \tag{2.8}$$

where x_i is the mole fraction of component i. $p_S(\sigma)$ is used to iteratively calculate the chemical potential of the surface segments, $\mu_S(\sigma)$ in the solution. $\mu_S(\sigma)$ is also called the "σ-potential".

$$\mu_S(\sigma) = -\frac{RT}{a_{\text{eff}}} \ln\left[\int p_S(\sigma') \exp\left\{\frac{a_{\text{eff}}}{RT}\left[\mu_S(\sigma') - E_{\text{misfit}}(\sigma,\sigma') - E_{HB}(\sigma,\sigma')\right]\right\} d\sigma'\right] \tag{2.9}$$

where a_{eff} is the effective contact area between two segments, $E_{\text{misfit}}(\sigma,\sigma')$ describes the interaction energy which is a result of the misfit between the SCDs, and $E_{HB}(\sigma,\sigma')$ describes hydrogen bonding. The σ-potential of the solution is finally used to calculate the chemical potentials, $\mu_S^{X_i}$, of the components in the solution.

$$\mu_S^{X_i} = \mu_{C,S}^{X_i} + \int p^{X_i}(\sigma)\mu_S(\sigma)d\sigma \tag{2.10}$$

where $\mu_{C,S}^{X_i}$ is a combinatorial contribution (Eckert and Klamt, 2002).

The COSMO work is available in a commercial software package named COSMO*therm*. Consequently a lot of details have not been published as they are company secrets. It can be considered a black box model where the input is the compound structures and the output is whatever predictions you desire. There exists many publications describing the application of the work in areas such as VLE calculations (Klamt and Eckert, 2004), activity coefficients at infinite dilution (Putnam *et al.*, 2003), LLE, SLE, Henry constants and vapor pressure (Eckert and Klamt, 2002).

The vapor pressure prediction of the COSMO method are generally unreliable both qualitatively and quantitatively (Eckert and Klamt, 2001). When experimental vapor pressures are applied the binary VLE predictions are generally less precise than UNIFAC (Eckert and Klamt, 2002), but the results can be considered good enough for initial design phases. Recent work by Sapei *et al.* (submitted) shows weaknesses for mixtures with sulfur containing compounds with respect to γ^{∞} and azeotropic properties. Currently there is not implemented any method for handling systems at elevated pressures and users of COSMO*therm* are forced to assume ideal gas behavior or combine it with an independent model for auxiliary phases.

The litterature of recent years is rich in articles covering bitter polemics over the results and potential of the theory behind the COSMO*therm* software (Wang *et al.*, 2006). The ongoing controversy, now longer than the besiege of Troy, makes it difficult to get an unbiased and justified view of COSMO. Certain articles tend to advertise rather than evaluate their results (Klamt *et al.*, 2002). Other articles claim difficulties in reproducing early results published by the founders, and seem to question every little detail in the early publications. The only way to explore COSMO*therm* in design decisions seems to be by personally gaining experience through long time use of the software. Thereby getting the knowledge of weaknesses and strength by trial and error. This is probably the biggest weakness of the approach.

2.3 Gibbs Free Energy Integration

Haile and Chialvo developed the "Mutiple-Parameter Charging Approach" where the excess Gibbs free energy is determined for a binary mixture through a set of isothermal-isobaric (NPT) molecular simulations (Haile, 1986; Chialvo and Haile, 1987; Chialvo, 1990). This approach is generally recommended for

accurate results. The main goal is to determine the excess Gibbs free energy, G^E:

$$G^E = G_{mix} - G_{is} \tag{2.11}$$

by integration. G_{mix} is the Gibbs energy of the mixture and G_{is} is the free energy of the ideal solution:

$$G_{is} = \sum x_i G_i^{pure} + kT \sum x_i \ln x_i \tag{2.12}$$

G_i^{pure} is the Gibbs free energy of pure i. For a binary mixture of compounds A and B the G^E can also be written (Haile, 1986):

$$G^E = x_A \left(G_{mix}^{res} - G_{pure\,A}^{res}\right) + x_B \left(G_{mix}^{res} - G_{pure\,B}^{res}\right) \tag{2.13}$$

where the residual energies are given by

$$G^{res} = G_{mix}(T,P,\mathbf{x}) - G_{ig}(T,P,\mathbf{x}) \tag{2.14}$$

with G_{ig} being the free energy of an ideal gas mixture at the same conditions as the mixture with respect to temperature, pressure and composition. G^{res} is related to the partition function function, Δ_C, of the isothermal-isobaric ensemble,

$$NG^{res}(x_A,T,P,\underline{\lambda}) = -kT \ln \Delta_C \tag{2.15}$$

where N is the number of molecules, x_A is the mole fraction of component A, T is the temperature and P is the pressure. $\underline{\lambda}$ is the parameter vector with k members applied in simulations, and Δ_C is given by

$$\Delta_C = C \int \exp\left(-\frac{PV}{NkT}\right) \partial V \int \ldots \int \exp\left(-\frac{U(\underline{r}^N,\underline{\omega}^N,\underline{\lambda})}{NkT}\right) \partial \underline{r}^N \partial \underline{\omega}^N \tag{2.16}$$

where $\underline{r}^N$ and $\underline{\omega}^N$ are the positions and angle orientations of the molecules. U is the intermolcular potential function.

The differential of the residual free energy is:

$$dG^{res} = \sum_{i=1}^{k} \left(\frac{\partial G^{res}}{\partial \lambda_i}\right) d\lambda_i \tag{2.17}$$

combining the expression above it is possible to derive the connection between dG^{res} and U can be derived

$$dG^{res} = \frac{1}{N} \sum_{i=1}^{k} \left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle d\lambda_i \tag{2.18}$$

where the angle brackets refers to an isothermal-isobaric ensemble average value:

$$\langle(\cdots)\rangle = \frac{\int \exp\left(-\frac{PV}{NkT}\right) \partial V \int \ldots \int \exp\left(-\frac{U(\underline{r}^N,\underline{\omega}^N,\underline{\lambda})}{NkT}\right) \partial \underline{r}^N \partial \underline{\omega}^N}{\Delta_C} \tag{2.19}$$

With the equation above the basis of the charging process is available. The simulation start with pure A or B and through a series of simulations the parameters of the component not present in the initial simulation is gradually charged until the parameter vector is equal to the parameter vector of the mixture. Thereby one retrieves the differences in the residual energies from Equation 2.13:

$$G^{res}_{mix} - G^{res}_{pure\,A} = \frac{1}{N}\sum_{i=1}^{k}\int_{\lambda_i^A}^{\lambda_i^{mix}} \left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle d\lambda_i \quad (2.20)$$

$$G^{res}_{mix} - G^{res}_{pure\,B} = \frac{1}{N}\sum_{i=1}^{k}\int_{\lambda_i^B}^{\lambda_i^{mix}} \left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle d\lambda_i \quad (2.21)$$

and the Gibbs excess energy can be calculated,

$$\begin{aligned} G^E = & \frac{x_a}{N}\sum_{i=1}^{k}\int_{\lambda_i^A}^{\lambda_i^{mix}} \left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle d\lambda_i \\ + & \frac{x_B}{N}\sum_{i=1}^{k}\int_{\lambda_i^B}^{\lambda_i^{mix}} \left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle d\lambda_i \end{aligned} \quad (2.22)$$

To evaluate the equation above one would need to make many simulations, however G^{res} is a state function and the value of ΔG^{res} is independent of the path of the integration of Equations 2.20 and 2.21. The shortest path through the k-dimensional space should be a straight line where the potential parameters are related through $k-1$ equations

$$\begin{aligned} c_2 &= \frac{\lambda_2^f - \lambda_2^i}{\lambda_1^f - \lambda_1^i} \\ c_3 &= \frac{\lambda_3^f - \lambda_3^i}{\lambda_1^f - \lambda_1^i} \\ &\vdots \\ c_k &= \frac{\lambda_k^f - \lambda_k^i}{\lambda_1^f - \lambda_1^i} \end{aligned}$$

where $c_1 = 1$ and the f and i superscript is final and initial values of the parameters. The diffential of G^{res} becomes

$$\begin{aligned} dG^{res} &= \frac{1}{N}\sum_{i=1}^{k}\left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle \left(\frac{\partial \lambda_i}{\partial \lambda_1}\right) d\lambda_1 \\ &= \frac{1}{N}\sum_{i=1}^{k}\left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle c_i d\lambda_1 \\ &= \frac{d\lambda_1}{N}\sum_{i=1}^{k} c_i \left\langle \left(\frac{\partial U}{\partial \lambda_i}\right)_{N,T,P,\lambda_{j\neq i}} \right\rangle \end{aligned} \quad (2.23)$$

Now the simulations have been reduced to two series of integrals, one from pure A and one from pure B to the mixture as shown in Figure 2.4.

The calculation of G^E can be expanded to calculation of the excess enthalpy, H^E, and the excess volume, V^E, (Chialvo, 1990). According to Chialvo and

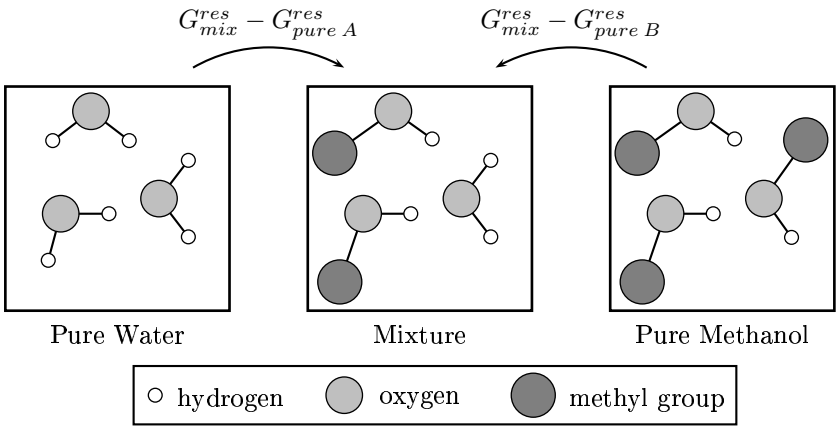


Figure 2.4. Visualization of GFEI for mixture of water and methanol.

Haile (1987); Chialvo (1990) the method is precise and secondly the standard deviation of the predictions is low. H^E and V^E can also be determine in the isothermal-isobaric ensemble by performing simulations of the pure components and then mixtures. The excess properties are then determined by finding the difference between the mixture value and the linear combination of the pure component results. The problem with that approach is the standard deviation which typically is larger than the excess value because the excess value is a small difference between two large numbers. This can be avoided by using the GFEI approach.

The obvious drawback of the method is the multiple simulations needed for each state point. More challenging is that simulation software has to be able to sample $(\partial U/\partial \lambda)$ to calculate the average value for Equation 2.23 for non-trivial force fields. The feature is not implemented in known free or commercial programs.

But, the major difficulty of the method is the problems when going from Lennard-Jones particles to real flexible molecules with UA or AA descriptions. The basic problem is expansion from particles and rigid molecules with similar topologies to flexible molecule combinations where simplifications can not be applied. The GFEI have been applied to simple particles (Chialvo and Haile, 1987) and to molecules of same topology (Slusher, 1998, 1999). Figure 2.4 show an example of a UA description applied to a water and methanol mixture thereby both molecules can be considered as three beads on a string and the GFEI is possible. The problem is when the topologies are no longer similar,

for example a water - ethanol mixture, where water consists of three beads as described before, but ethanol consists of four beads. The GFEI theory does not allow growing of additional beads. This reduces the application area to special cases.

If the GFEI method is going to be competitive the theoretical derivation for handling components of different topology has to be created and a user friendly software package need to be available which can handle the sampling of $(\partial U/\partial \lambda)$. If precision and standard deviation of the prediction lives up to the claims, the method can have a future in the set of general molecular modeling tools applied by chemical engineers.

2.4 Summation

The three methods presented all have different strengths and weaknesses from the viewpoint of a chemical engineer. It is the author's opinion that there is still a lack of a method which is accurate and easy to use. The results should be obtainable without the user has expert knowledge of setting up and analyzing simulations. The method should also be available through ready-to-use cheap software.

3

Fluctuation Solution Theory

The fluctuation solution theory (FST) gives the statistic mechanical connection between the angle-averaged radial distribution function (RDF) and the chemical potential, μ, through the grand canonical ensemble, GCE or μVT.

$$\left(\frac{\partial \mu_1}{\partial x_1}\right)_{T,P} = kT\left(\frac{1}{x_1} - \frac{c_2\left(G_{11}+G_{22}-2G_{12}\right)}{1+x_1c_2\left(G_{11}+G_{22}-2G_{12}\right)}\right) \tag{3.1}$$

where G_{ij} is the volume integral of RDF from the grand canonical ensemble.

$$G_{ij} = \int_0^\infty \left(g_{ij}^{\mu VT}(r) - 1\right) r^2 dr \tag{3.2}$$

The theory is known as the statistical mechanical theory of solutions (Kirkwood and Buff, 1951) and the comprehensive derivation is provided in Appendix A. The theoretical development was continued by O'Connell (1971a,b), who derived the key expression applied in this work using the Total Correlation Function Integrals, TCFIs or H_{ij}'s:

$$\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)_{T,P} = -\frac{x_2 \Delta H}{1+x_1x_2\Delta H} \tag{3.3}$$

$$\Delta H = H_{11} + H_{22} - 2H_{12} \tag{3.4}$$

$$H_{ij} = \rho G_{ij} \tag{3.5}$$

The TCFIs can also be determined from a set of properties using a reverse approach (Ben-Naim, 1977; Matteoli and Lepori, 1984; Wooley and O'Connell, 1991). Wooley and O'Connell (1991) generated a data bank of TCFIs for binary mixtures using such a reverse approach. Experimental data was applied in the regression of parameters for various correlations where Redlich/Kister-like polynomials (Handa and Benson, 1979) were used for densities, ρ, and partial molar volumes, $\bar{V}_i$ and the Huang-O'Connell model for isothermal compressibilities, κ_T (Huang and O'Connell, 1987). Finally the NRTL, modified Margules or Wilson G^E-models from phase equlibria used to calculate activity coefficient derivatives, $(\partial \ln \gamma_1/\partial x_1)_{T,P}$ (Abbott and van Ness, 1975). The route to TCFIs was via integrals of the direct correlation function of Ornstein and Zernike, C_{ij},

$$(1 - C_{11}) = \frac{\rho \bar{V}_1^2}{\kappa_T RT} + x_2\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)_{T,P} \tag{3.6}$$

$$(1 - C_{12}) = \frac{\rho \bar{V}_1 \bar{V}_2}{\kappa_T RT} - x_1 \left(\frac{\partial \ln \gamma_1}{\partial x_1} \right)_{T,P} \tag{3.7}$$

$$(1 - C_{22}) = \frac{\rho \bar{V}_2^2}{\kappa_T RT} + x_1 \left(\frac{\partial \ln \gamma_2}{\partial x_2} \right)_{T,P} \tag{3.8}$$

which were then converted to TCFIs by an integrated form of the Ornstein-Zernike equation (O'Connell, 1971a),

$$H_{11} = \left(\frac{1 - x_2 C_{22}}{D} - 1 \right) \frac{1}{x_1} = \frac{C_{11} - x_2 \left(C_{11} C_{22} - C_{12}^2 \right)}{D} \tag{3.9}$$

$$H_{12} = \frac{C_{12}}{D} \tag{3.10}$$

$$H_{22} = \left(\frac{1 - x_1 C_{11}}{D} - 1 \right) \frac{1}{x_2} = \frac{C_{22} - x_1 \left(C_{11} C_{22} - C_{12}^2 \right)}{D} \tag{3.11}$$

$$D = (1 - x_1 C_{11})(1 - x_2 C_{22}) - x_1 x_2 C_{12}^2 \tag{3.12}$$

Through the reverse approach it is possible to evaluate the individual TCFI's (H_{11}, H_{12} or, H_{22}) when studying a binary system which will be applied in most of the case studies in Chapter 7. In the rest of the thesis $(\partial \ln \gamma_1 / \partial x_1)_{T,P}$ is simplified to $d \ln \gamma_1 / dx_1$.

3.1 Practical Fluctuation Solution Theory

Previous simulation work suggests that FST analysis of simulation results can provide quantitative information concerning the thermodynamics of solutions (Matteoli and Mansoori, 1990; Weerasinghe and Smith, 2003). In this chapter the practical approach for generation of TCFIs is presented where the basic element is isothermal-isobaric molecular dynamic (NPT-MD) simulations of liquid mixtures at different mixing ratios. The NPT ensemble is preferred rather than the grand canonical ensemble since it is more convenient in MD simulations. Also the problems with respect to insertion moves are avoided. The details are described in section 3.2 together with a short introduction to MD simulations and its opportunities and restrictions. The radial distribution functions are the product of the simulations and in Section 3.3 the treatment of the RDF and their integration to get TCFIs is described

3.2 MD simulations

Molecular dynamics is computer simulations where atoms or molecules interact using Newtonian mechanics and integrating over time (Haile, 1997). By sampling the simulations it is possible to get a time average of the properties of interest. In our case the most interesting properties are the molecule positions and the volume. The Newtonian mechanics are represented by a force field

which in this work is the all-atom CHARMM described later in detail. An AA description was selected because it has been assumed that a simpler UA description would affect the RDF. The application of a AA force field increases the requirements for the software and computer power.The NAMD software package of Kalé *et al.* (1999) was taken into use on the cluster Horseshue of the Danish Center for Scientific Computing at the University of Southern Denmark. NAMD has several advantages which includes the ability to exploit multiprocessor computers and multiple computers in a cluster to speed up calculations. Of similar importance, it was developed and optimized for MD simulation applying the CHARMM force field. Periodic boundary conditions were imposed, and the time step used in the integration was 1fs. A cut-off distance of 12A was applied together with a switching function at 10A. The particle-mesh-Ewald (PME) algorithm was used to calculate long-range electrostatic forces Kalé *et al.* (1999).

The starting configuration was built from the equilibrium configuration of a single molecule. The simulation box was divided into a simple grid and the molecule templates were inserted at the grid points in accordance with the desired composition. System sizes of 512 and 1000 molecules have been used. Two steps were used for equilibrating the system. Firstly, the system is relaxed by a steepest descent method to eliminate atomic overlaps in the initial setup of the system. This process was terminated when the system potential energy became essentially constant. Secondly, the NPT simulations (pressure of 1 atm and a temperature equal to the experimental value) were started and ran for 500ps. The criterion for terminating at this stage was that there was negligible variation in the radial pair distribution functions. Then the production period was started. Its length varied somewhat with the composition. In most cases, 3-4 production periods of 5-6 ns were used to obtain average and variance of TCFIs.

3.2.1 Force Field

The all-atom CHARMM force field has mainly been applied in the study of biomolecules (A. D. MacKerell *et al.*, 1998; A. D. MacKerell and Banavali, 2000), however Liu *et al.* (2004) use it for two halogenated compounds, haloethane and hexafluoroethane, and Chen *et al.* (2002) studied fluoro ethanes. The force field is characterized by harmonic potentials for flexible bonds, a Urey-Bradley (UB) term for 1-3 interactions and a Lennard-Jones term for 1-4 interactions and other non-bonded interactions in general. The UB term was applied consistently and in special cases a modified parameter set was used for 1-4 L-J interactions (A. D. MacKerell *et al.*, 1998). Coulombic forces are included using charges positioned at the centers of mass of the atoms where interactions are included for atoms separated by three or more bonds. The general form of the potential is given by:

$$U(R) = \sum_{bonds} K_b \left(b - b_0\right)^2 + \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2 + \sum_{UB} K_{UB} \left(S - S_0\right)^2$$

$$+ \sum_{dihedrals} K_\chi \left(1 + \cos\left(n\chi - \delta\right)\right) + \sum_{impropers} K_\varphi \left(\varphi - \varphi_0\right)^2$$
$$+ \sum_{nonbond} \left(\epsilon_{ij} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{\epsilon_1 r_{ij}} \right) \quad (3.13)$$

where K_b, K_{UB}, K_θ, K_χ, and K_φ are the bond, Urey-Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; b, S, θ, χ, and φ are the bond length, Urey-Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, and the subscript zero represents equilibrium values for the individual terms. n is the symmetric number number and δ is the phase shift. ϵ_{ij} is the Lennard-Jones well depth and $R_{min,ij}$ is the distance at the minimum of the Lennard-Jones potential. The unlike Lennard-Jones parameters are calculated with Lorentz-Berthelot combining rules. q_i is the atomic partial charge, ϵ_1 is the effective dielectric constant and r_{ij} is the distance between atoms i and j. The dielectric constant is set to one in all calculations, corresponding to the permittivity of vacuum (Foloppe and Mackerell, 2000). The parameter tables for the different systems are listed in Appendix C.

The setup of input files for the simulations are straightforward and only basic knowledge of MD simulations are needed to before a novice can start making MD simulations. This easy accessibility of MD is quit opposite to advanced forms of MC like GEMC and GCMC, where one can fear improper sampling of the system caused by incorrect sampling of the configurational space as mentioned in Section 2.1.

3.2.2 MD versus MC

The reasons for selection of MD over MC needs to be established. Mainly because the time integration in MD is computationally expensive compared to the random moves in MC simulation. When only equilibrium or thermodynamic properties, A, are studied the ergodic hypothesis states that the statistical average of MC simulation is equal to the time average of MD simulation.

$$\langle A \rangle_{\mathrm{MC}} = \langle A \rangle_{\mathrm{MD}} \quad (3.14)$$

Therefore the results of comparative studies should be the same. From this simplified point of view one should choose MC over MD, but several challenges are introduced when AA force fields are applied, even to small organic molecules, for example n-hexane or benzene.

In section 2.1 it was described how it initially was impossible to partially or completely move a UA molecule into a dense phase (Panagiotopoulos *et al.*, 1988). The problem was solved by the Configurational-bias Monte Carlo strategy (Laso *et al.*, 1992; Vlugt *et al.*, 1992; Martin and Siepmann, 1999). However the CBMC strategy was developed for AU force fields where bond length distributions are not taken into consideration and more importantly the hydrogen

atoms not included. Martin and Frischknecht (2006) have recently developed a new formulation of configurational-bias Monte Carlo strategy which aims at improved sampling of the intramolecular configurational phase space. The problem still remains that even if the best trial state is selected for insertion of parts or a complete molecule the chance of the insertion is the product of chances of each bond length, bond angle, dihedral angle and so forth which makes correct sampling of the configurational space challenging when applying an AA force field description. Consequently very long equilibration and production periods are needed. This problem is not encountered in MD simulations where the configurational space is automatically correctly covered through the time integration of the Newton equations.

The development of software has also limited the advantages of MC over MD with software packages like NAMD (Kalé *et al.*, 1999). NAMD has the ability to run on parallel computers in a cluster and multiple cpu's with multiple cores on each node or computer in the cluster. These impressive characteristics makes it possible to simulate large system with AA descriptions at a reasonable time scale. Because of the nature of MC it is not possible to explore the power of multiple computers in a cluster but the Towhee[1] software package has the capability of exploiting a computer with multiple cpu's like dual pentium. With system sizes of 512 or 1000 molecules the MC simulations on a new desktop computer would take weeks for each composition while the MD simulations could be finished within 7-10 days on a cluster depending on the capacity available.

3.3 RDF from Simulation and their Integration; 3 Methods

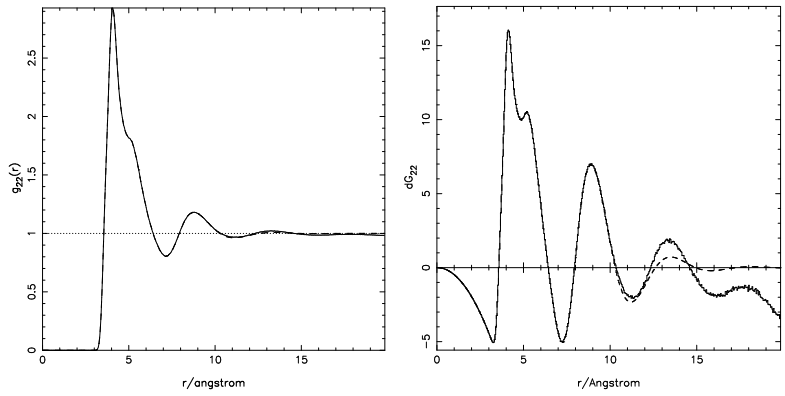
After positions of the molecules have been sampled in the MD simulations the radial distribution functions, $g_{ij}^{NPT}(r)$, can be determined and integrated to retrieve the TCFIs.

$$H_{ij} = 4\pi\rho \int_0^{r_{max}} r^2 \left(g_{ij}^{NPT}(r) - 1\right) dr \tag{3.15}$$

The integral in Equation (3.15), has a finite upper limit, r_{max}. The treatment of this can be complicated by the limited range, set by the simulation box, for which we may simulate $g_{ij}(r)$, and the slow convergence of $g_{ij}(r)$ toward the limiting value of unity. The maximum distance, r_{max}, equals half the side length of the simulation box, L_{box}. Theoretically, $g_{ij}(r) - 1$ goes to zero when r goes to infinity. However, because the integral is evaluated numerically, convergence requires that $g_{ij}(r)-1$ goes faster to zero than r^2 goes to infinity. If not, the integral diverges. Occasionally, this can create problems. For example, convergence of $g_{ij}(r)$ may not be attained at $r = r_{max}$. The deviation of

[1] http://towhee.sourceforge.net

$g_{ij}(r)$ from unity may be about $0.1 - 0.2\%$ or less. Yet, this magnifies in the TCFI because of the r^2 factor in the integrand, as illustrated in Figure 3.1 for the ethanol-ethanol RDF for the benzene-ethanol system which is treated later in Section 7.3. The solid line in Figure 3.1(a) is $g_{22}(r)$ which visually does not deviate from unity near r_{max}. However, the divergence is clearly seen on Figure 3.1(b) starting at $r \approx 15A$ where a local maximum in the integrand, dG_{22}, has a negative value. To overcome these difficulties three levels of approximation have been considered. The first approach (Method 1) is



(a) $g_{22}(r)$ at $x_1 = 0.5$ (b) Integrand, dG_{22} at $x_1 = 0.5$

Figure 3.1. Effect of divergence of $g_{ij}(r)$ at longer distances. $x_1 = 0.5$, where 1: benzene and 2: ethanol. Solid line represents Method 1. Dashed line is a fit using Equation (3.21) with Method 3.

simply to stop integration when a sub unity maximum or above unity minimum in $g_{ij}(r)$ is encountered thereby reducing r_{max}. This gives an approximate H_{ij} value. The determination of the reduced value of r_{max} has proven to be a challenge because the RDF can be influenced by less perfect sampling at large distance close to the simulation box length. Consequently a simple but robust algorithm has been developed and implemented to find possible divergence and to thoroughly test the RDF in the vicinity of the detected divergence to ensure its correctness. Method 2 is to increase the size of the system simulated and redo the simulations. Literature provides examples of this (Weerasinghe and Smith, 2003), although it increases the simulation time significantly. In the case study of benzene - methyl acetate (Sec. 7.1) the system size was increased from 512 to 1000 molecules and the integrand from this system still diverge when approaching the half simulation box length. However as the box size has been increased the distance of divergence has also been increased, $r_{max}^{1000} > r_{max}^{512}$, which increases the precision of H_{ij}. Method 3 uses some expression to reproduce the RDF which approaches unity faster than r^2 goes to infinity. The methods have been summarized below.

1. Set r_{max} as the limit where RDF starts to diverge.
2. Increase number of molecules and repeat simulations.
3. Expression to reproduce RDF which guarantees convergence.

In the literature examples can be found of expressions applied for reproduction of the RDF. Matteoli and Mansoori (1995) proposed an expression for reproducing the RDF of mainly pure Lennard-Jones fluids which also included an extension to binary mixtures. The expression contains two equations, one to represent the behavior before the maximum of the first peak and a second equation to represent the rest. It is assumed that the RDF is continuous and the first maximum occur at the contact distance, d. d is given by $d = h\sigma$, where σ is the distance parameter of the Lennard-Jones potential and h is an adjustable parameter near unity.

$$g(y) = \begin{cases} 1 + y^{-m}\left(g(d) - 1 - \lambda\right) & \\ +\frac{y-1+\lambda}{y}\exp\left(-\alpha(y-1)\right)\cos\left(\beta(y-1)\right) & m \geq 1,\, y \geq 1 \\ g(d)\exp\left(-\theta(y-1)^2\right) & y < 1 \end{cases} \tag{3.16}$$

The reduced radius, y, is used to represent the distance which is given by $y = r/d$. The adjustable parameters are $h = d/\sigma$, m, λ, α, β, θ and $g(d)$ which is the height of the first peak. The function also assumes that the first positive peak is the maximum as the function is a decay with respect to the radius. The expression (Eq. 3.16) can be modified to make it applicable to binary mixtures. The contact distances, σ_A and σ_A, are independent of mole fraction (3.17), but they vary among g_{AA}, g_{AB} and g_{BB} and the three h_{ij}'s describe the difference.

$$d = h_{ij}\frac{\sigma_A + \sigma_B}{2} \tag{3.17}$$

A new dependent variable, Δr_M, is introduced (3.18) which the distance between a local maximum and the subsequent local maximum.

$$\Delta r_M = h\left(X_{VA}\sigma_A + X_{VB}\sigma_B\right) \tag{3.18}$$

where X_{Vi} is the volume fraction which given by:

$$X_{VA} \cong \frac{x_A\sigma_A^3}{x_A\sigma_A^3 + x_B\sigma_B^3} \tag{3.19}$$

With the equations above it is possible to expand the expression for the reduced radius, y.

$$y = \begin{cases} \frac{r}{d} & r < d \\ \frac{r-d+\Delta r_M}{\Delta r_M} & r \geq d \end{cases} \tag{3.20}$$

With the set of equations above it should be possible to represent g_{ij} for all compositions. However, the procedure presented by Matteoli and Mansoori

(1995) has a number of disadvantages and limitations. From a mathematical point of view the expression becomes quite complex and from numerical point of view it becomes difficult to determine the adjustable parameters even when good initial estimates are available.

A RDF can be divided into two parts, direct and indirect interaction, where the direct or first part describes the interactions of a molecule and the molecules next to it which are bumping into each other. The direct interaction part starts at $r = 0$ and goes onto the third unity of $g(r)$ (see Figure 3.1.a). The indirect part of $g(r)$ is everything beyond the third unity where there is no direct interaction between the molecules. The expression of Matteoli and Mansoori (1995) includes only a single peak for the direct interaction part of the RDF which is acceptable for L-J fluids. However, the direct interaction part can become a combination of several peaks when describing real liquids which is seen in Figure 3.1.a for ethanol. As a consequence, the procedure of Matteoli and Mansoori (1995) can not be employed in its current form, without fundamentally new developments.

The direct interaction part of the RDFs from MD simulations were considered to be correct as convergence problems start at larger distances. The numerical integration of the direct interaction part of the RDF was considered correct, thereby only the indirect interaction part of the RDF has to be reproduced by the expression.

The idea of combining a declining exponential and a wave function to reproduce the the indirect interaction part of the RDF is valid for real molecules as they at longer distances behave like hard spheres, but the complicated construction of the expression by Matteoli and Mansoori (1995) made it impossible to apply. A simple expression has been developed to reproduce the indirect interaction part:

$$g_{ij}^{indirect}(r) = 1 + a \exp\left(b(r-c)\right) \sin\left(d(r-e)\right) \tag{3.21}$$

The 5 parameters are determined by regression where the objective is to reproduce $g_{ij}(r)$ from the third unity, r_{u3}, until either r_{max} or the distance where $g_{ij}(r)$ diverges. An example is shown in Figure 3.1 where the dashed curve is the result of such a regression. When applying the expression it is possible to integrate the $g_{ij}^{indirect}$ until r_{max}. But it is also possible to extrapolate using the expression from r_{max} to radius, r_{nc}, where there is no longer any contribution to the TCFI. The contribution to H_{ij} from integration from r_{max} to r_{nc} is viewed as a long distance correction term, H_{ij}^{ld}, and practically this approach enables integration from zero to infinity. The final form for TCFIs has three contributions. The direct interaction part of the RDF is integrated numerically giving first contribution, H_{ij}^{direct}. The second contribution is the integration of Equation 3.21 from the start of the indirect interaction part of the RDF (third unity) to the maximum distance determined by the box size, H_{ij}^{box}, and the final term is the long distance correction term, H_{ij}^{ld}.

$$H_{ij} = H_{ij}^{direct} + H_{ij}^{box} + H_{ij}^{ld} \tag{3.22}$$

The statistical uncertainty of the contribution from the direct interaction part of the RDFs, H_{ij}^{direct}, was negligible. The typically standard deviation of H_{ij}^{direct} is in the range of 0.01, but the standard deviation like-like direct integral, $\sigma_{H_{ii}^{direct}}$, is composition dependent. $\sigma_{H_{ii}^{direct}}$ ranges from 0.001 at high concentrations of i to 0.2 at dilute concentrations. $\sigma_{H_{ij}^{box}}$ from the correlation (Eq. 3.21) is averagely a factor 5 to 10 lower than $\sigma_{H_{ij}^{direct}}$. The long distance contribution, H_{ij}^{ld}, is the smallest of the three contribution and insignificant in size. H_{ij}^{ld} is generally smaller than the standard deviation of the TCFI, H_{ij}, and it makes up for $1-3\%$ of the TCFI. There are examples where the H_{ij}^{direct} and H_{ij}^{box} almost cancel out each other, thereby H_{ij}^{ld} makes up for a relative big part of the TCFI, however this does not change the overall result which is a derivative, $\partial \ln \gamma_1 / \partial x_1$, close to zero.

In some special cases it is not possible to use Method 3 due to the shape of $g_{ij}(r)$. In these cases Methods 1 or 2 should be applied. In general Method 1 has been preferred in this work.

4

Generation of G^E-model Parameters

The next step in the procedure is to explore the information about the macroscopic behavior of the mixtures to generate a descriptive model covering the whole composition range. One could integrate the derivatives of the activity coefficients to get the activity coefficient curves as function of composition, however the number of compositions needed to be simulated would be enormous. Instead a small set of compositions are simulated and a G^E-correlation is used to reproduce the derivatives gained from the simulations. This formulation uses the Lewis/Randall pure-component standard state. Note that the original treatment of Kirkwood and Buff (1951) used the Henry's Law standard state, whereas a later development by O'Connell (1971a) allowed a choice of standard state. A likely advantage of building activity coefficient models from derivatives is that the ultimate properties are less sensitive to errors in the basic model.

The modified Margules model has been chosen in this work (Abbott and van Ness, 1975),

$$\frac{G^E}{RTx_1x_2} = A_{21}x_1 + A_{12}x_2 - \frac{\alpha_{12}\alpha_{21}x_1x_2}{\alpha_{12}x_1 + \alpha_{21}x_2 + \eta x_1x_2} \tag{4.1}$$

The expression is used in three forms: The simplest form is sufficient for nearly ideal systems and only includes A_{12} and A_{21}. For non-ideal systems extension to either 4 parameters (A_{ij}'s and α_{ij}'s) or all 5 parameters is normal.

4.1 Optimization and Objective Functions

The model parameters for the simulations were determined by minimizing an objective function. In the most of the Ph.D. work a simple objective function, SS_1, was applied (Christensen *et al.*, 2007, in pressb),

$$SS_1 = \sum_i \left[\left(\frac{\partial \ln \gamma_1}{\partial x_1} \right)_{MD,i} - \left(\frac{\partial \ln \gamma_1}{\partial x_1} \right)_{mM,i} \right]^2 \tag{4.2}$$

Subscript MD denotes results obtained from FST analysis of MD results (Eq. 3.3). mM denotes derivatives calculated by the modified Margules model using the adjustable parameters which is obtained by taking the derivative of Equation 4.1 with respect to x_1 twice. The summation is over all simulated compositions with one term for each production period.

Uncertainties are typically encountered when one component is diluted, giving a less reliable like-like RDF. This phenomenon is particularly pronounced for systems close to ideal behavior when ΔH is small and sensitive to variances of H_{ii}'s. A new objective function, SS_2, has been investigated to overcome the difficulties (Christensen *et al.*, in pressa),

$$SS_2 = \sum_i \frac{\left[\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)_{MD,i}^{avg} - \left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)_{mM,i}\right]^2}{\sigma_i^2 \left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)_{MD}} \tag{4.3}$$

where the superscript avg denotes the average derivative at composition i over the production periods. σ_i^2 is the variance of derivative from the MD simulations. The result of MD simulations is not the derivative,

$$\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)_{MD}^{avg} = \frac{-x_2\left(H_{11}^{avg} + H_{22}^{avg} - 2H_{12}^{avg}\right)}{1 + x_1 x_2 \left(H_{11}^{avg} + H_{22}^{avg} - 2H_{12}^{avg}\right)} \tag{4.4}$$

but the TCFI's and their averages and variances. The variance of the derivative

$$\sigma_i^2\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)_{MD} = \sigma_{H_{11}}^2\left(\frac{\partial\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)}{\partial H_{11}}\right)^2 + \sigma_{H_{12}}^2\left(\frac{\partial\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)}{\partial H_{12}}\right)^2 + \sigma_{H_{22}}^2\left(\frac{\partial\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)}{\partial H_{22}}\right)^2 \tag{4.5}$$

is a combination of the variances of the TCFIs. In this way, results with an uncertain $(\partial \ln \gamma_1/\partial x_1)_{MD,i}$ has less weight, and thereby less relative impact on the objective function, SS_2, compared to SS_1. The $\sigma_{H_{ij}}^2$'s are output from the analysis of the MD simulations and the derivatives with respect to H_{ij}'s are obtained by deriving Equation 3.3.

$$\frac{\partial\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)}{\partial H_{11}} = \frac{-x_2}{\left(1 + x_1 x_2 \Delta H\right)^2} \tag{4.6}$$

$$\frac{\partial\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)}{\partial H_{12}} = \frac{2x_2}{\left(1 + x_1 x_2 \Delta H\right)^2} \tag{4.7}$$

$$\frac{\partial\left(\frac{\partial \ln \gamma_1}{\partial x_1}\right)}{\partial H_{22}} = \frac{-x_2}{\left(1 + x_1 x_2 \Delta H\right)^2} \tag{4.8}$$

The traditional optimization procedure Abbott and van Ness (1975) has 3 steps where an initial optimization only includes 2 parameters (A_{12} and A_{21}). This is

followed by a 4- parameter optimization adjusting A_{12}, A_{21}, α_{12} and α_{21}. If the goodness-of-fit increases sufficiently, the 4 parameter set is temporarily chosen. A final 5-parameter minimization is made adjusting also η. If again goodness-of-fit increases the 5 parameter set is selected. A more systematic procedure uses the parameter variances resulting from minimizing SS_1 or SS_2. From each minimization an estimate can be made of the variance of the resulting parameters.

A good indication of choosing a simpler modified Margules expression is when the order magnitude of one or more σ_{a_j}'s resembles that of the a_j's. Even though the value of the objective function may increase slightly, the simpler expression should be chosen for the final parameter set.

4.2 Variance of Parameters

The variances of the parameters $\mathbf{a}$, $\sigma^2_{a_j}$, are calculated from the variance-covariance matrix of the underlying minimization problem.

4.2.1 General Newton-Raphson method

For solving non-linear minimization problems a general approach has been proposed called the Newton-Raphson method. The starting point of explaining the basic method is the objective function which is a squared sum of residuals and two examples are applied in this work (Equations 4.2 and 4.3). The objective function, SS, can be on given two forms:

$$\begin{aligned} SS &= \sum_{i=1}^{m} r_i^2 \\ &= \mathbf{r}^T\mathbf{r} \end{aligned}$$

where r_i are the individual residuals in the residual vector, $\mathbf{r}$, where the length is the number of residuals, m. The gradient of the objective function, ∇SS, with respect to the parameters are given by:

$$\nabla SS = \begin{bmatrix} \frac{dSS}{da_1} \\ \frac{dSS}{da_2} \\ \vdots \\ \frac{dSS}{da_i} \\ \vdots \\ \frac{dSS}{da_n} \end{bmatrix} \tag{4.9}$$

where a_i are the parameters of the parameter vector $\mathbf{a}$ which has n members. The second derivatives of the objective function, $\nabla^2 SS$, with respect to the

parameters is called the Hessian matrix, $\mathbf{H}$:

$$\nabla^2 SS = \mathbf{H} = \begin{bmatrix} \frac{d^2SS}{da_1^2} & \frac{d^2SS}{da_1 da_2} & \cdots & \frac{d^2SS}{da_1 da_n} \\ \frac{d^2SS}{da_2 da_1} & \frac{d^2SS}{da_2^2} & \cdots & \frac{d^2SS}{da_2 da_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d^2SS}{da_n da_1} & \frac{d^2SS}{da_n da_2} & \cdots & \frac{d^2SS}{da_n^2} \end{bmatrix} \tag{4.10}$$

Two criteria must be satisfied before a local minimum has been found. First the gradient must only contain zeros which is the reason why the method is also called zeroing the gradient. Second the Hessian matrix is positive definite. The second criterion means that the solution in fact is a local minimum. An example where the first criteria is satisfied but not the second is in the case of a saddle point.

There are 3 basic steps in each iteration when solving a minimization using the general Newton-Raphson method:

1. Evaluate the gradient $\nabla SS\left(\mathbf{a}^i\right)$ and the Hessian matrix $\mathbf{H}\left(\mathbf{a}^i\right)$.
2. Solve linear system $\mathbf{H}\left(\mathbf{a}^i\right)\mathbf{s}^i = -\nabla SS\left(\mathbf{a}^i\right)$. Where $\mathbf{s}^i$ is the change in the parameters to obtain the zero gradient at iteration i.
3. Calculate new estimated solution $\mathbf{a}^{i+1} = \mathbf{a}^i + \mathbf{s}^i$.

The general approach can be inefficient, especially when the initial estimate of the solution is bad.

4.2.2 Marquardt Method

The Marquardt method is an elegant algorithm which varies between the extremes of the inverse Hessian matrix (described above) and the steepest descent method. The method is very practical in the minimization of sum of squares for non-linear systems. The essential part of the algorithm is the damping factor, λ, which corrects the diagonal of the Hessian matrix. An inner loop is included in the Newton-Raphson algorithm where the damping factor is varied linearly to find the best $\mathbf{s}^i$ at $\mathbf{a}^i$ to give the lowest value of the objective function. The inner loop consists of the following steps:

1. Calculate $SS(\mathbf{a}^i)$.
2. Pick initial modest value of λ, for example $\lambda = 0.001$
3. Solve equation system to find $\mathbf{s}^i$, and calculate $SS(\mathbf{a}^i + \mathbf{s}^i)$.
4. If $SS(\mathbf{a}^i + \mathbf{s}^i) \geq SS(\mathbf{a}^i)$ increase λ by a factor, for example 10, and repeat step 3. Elseif $SS(\mathbf{a}^i + \mathbf{s}^i) < SS(\mathbf{a}^i)$ decrease λ and repeat step 3.

The inner loop is continued until the λ giving the lowest $SS(\mathbf{a}^i + \mathbf{s}^i)$ is found.

Both the general Newton-Raphson and the Marquardt method need the second derivatives for the Hessian matrix which are very comprehensive if not impossible to derive and calculate in many cases.

4.2.3 Modified Marquardt Method

The optimization algorithm applied is called the modified Marquardt method which is not as computationally heavy as the previously algorithms. The main reduction is made by introducing an approximate Hessian matrix where the second derivative of the residuals are not calculated. Another computational reduction is made in the inner loop where an expected change of the objective function is applied. To understand the changes the best starting point is the gradient of the objective function, ∇SS, where each element, l is given by:

$$\begin{aligned} \nabla SS_l &= \frac{dSS}{da_l} \\ &= \frac{d}{da_l}\left(\sum_{i=1}^{m} r_i^2\right) \\ &= 2\sum_{i=1}^{m} r_i \frac{dr_i}{da_l} \end{aligned} \tag{4.11}$$

The gradient can also be given in matrix form where dr_i/da_n are put in a $n \times m$ Jacobian matrix, $\mathbf{J}$.

$$\nabla SS = 2\mathbf{Jr} \tag{4.12}$$

The elements of the Hessian matrix are given by:

$$\begin{aligned} \mathbf{H}_{lk} &= \frac{d}{da_k}\left(2\sum_{i=1}^{m} r_i \frac{dr_i}{da_l}\right) \\ &= 2\left(\sum_{i=1}^{m} \frac{dr_i}{da_l}\frac{dr_i}{da_k} + \sum_{i=1}^{m} r_i \frac{d^2 r_i}{da_l da_k}\right) \\ &= 2\left(A_{lk} + \sum_{i=1}^{m} r_i \frac{d^2 r_i}{da_l da_k}\right) \end{aligned} \tag{4.13}$$

where A_{lk} is an element in the approximate Hessian matrix, A, which is applied in the modified Marquardt method instead of the real Hessian matrix (Eq. 4.10). A is calculated using the Jacobian matrix:

$$\mathbf{A} = \mathbf{J}\mathbf{J}^T \tag{4.14}$$

By applying $\mathbf{A}$ instead of $\mathbf{H}$ the number of calculations is reduced considerably and the need for taking second derivatives of r_i ($d^2 r_i/da_l da_k$) is eliminated. In the modified Marquardt algorithm the change in the parameters, $\mathbf{s}^i$, is determined by solving the system:

$$\left(\mathbf{A}^i + \lambda \mathbf{D}\right)\mathbf{s}^i = -\nabla SS \tag{4.15}$$

D is a diagonal matrix where the elements are either equal to the diagonal elements of **A** or a unit matrix. In the inner loop, where λ is determined, an expected change in the objective function, DQ, is introduced.

$$DQ = \left(2\nabla SS + \left(\mathrm{A}^i + \lambda \mathbf{D}\right)\mathbf{s}^i\right)\mathbf{s}^i \tag{4.16}$$

which is applied instead of actually solving $\mathbf{H}\left(\mathbf{a}^i,\lambda\right)\mathbf{s}^i = -\nabla SS\left(\mathbf{a}^i\right)$ and calculating the objective function with the new parameter set for the each λ-value. DQ is obviously less precise than the ordinary Marquardt method but the number of calculations is greatly reduced, especially in cases with many data point/residuals and parameters.

4.2.4 Co-variance Matrix and uncertainties

The variance-covariance matrix, **C**, is found from the approximate Hessian matrix of the final parameter set.

$$\mathbf{C} = \sigma^2(\mathbf{A}\mathbf{A}^T)^{-1} \tag{4.17}$$

An estimator for σ^2 is $SS/(m-n)$, where SS finanl value of the objective function. The diagonal elements of **C** give the variances of **a**.

$$\sigma_{a_j}^2 = C_{jj} \tag{4.18}$$

The variances of the parameters, $\sigma_{a_j}^2$, are converted to standard deviations, σ_{a_j}, listed in the parameter tables in the different case studies.

5

Prediction of Bubble Point Pressure

The modified Margules description generated by the FST methodology is applied in prediction of isothermal bubble point pressure curve of the binary mixture simulated. The G^E-model describes the non-ideality of the liquid mixture and a Poynting factor can be implemented to handle possible pressure dependence. The vapor phase behavior has not been characterized or analyzed through simulations. In this chapter several approaches for handling non-ideal behavior in vapor phase is considered. The vapor phase can be treated as either an ideal gas, ideal solution or real solution. The simplest approach is to assume ideality of the vapor phase however this is only reliable at low and ambient pressures. Ideal solution can be sufficient at elevated pressures when the unlike interactions can be assumed to be an average of the like-like interactions. If this assumption is not sufficient the vapor phase can be treated as a real solution using some equation of state.

The pressure predictions are the evaluation of the FST methodology presented in this thesis. A simple score is applied to evaluate each system which gives an average relative error of the pressure predictions

$$SCORE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{P_{exp} - P_{calc}}{P_{exp}} \right| \tag{5.1}$$

where n is the number of experimental data points, P_{exp} is the experimental pressure and P_{calc} is the calculated pressure. The score tables can be viewed in the different case studies in chapter 7.

Another important subject is the uncertainty of the pressure predictions which are determined by combining the rule of error propagation and the uncertainties of the model parameters from the previously chapter.

5.1 Treatment of the Vapor Phase

The objective is to solve the equations of equal fugacity for vapor-liquid equilibria,

$$y_i P \Phi_i = x_i \gamma_i P_i^{sat};\ i = 1{,}2 \tag{5.2}$$

where y_i is the mole fraction in the vapor phase, P is the pressure, x_i is the mole fraction in the liquid phase, γ_i is the activity coefficient, P_i^{sat} is the pure component vapor pressure at the specified temperature and Φ_i is the fugacity factor giving by:

$$\Phi_i = \frac{\hat{\phi}_i}{\phi_i^{sat} \mathrm{PF}_i} \tag{5.3}$$

$\hat{\phi}_i$ is fugacity coefficient in the vapor phase, ϕ_i^{sat} is the pure component fugacity coefficient at the saturation point, (T,P_i^{sat}), and PF_i is the Poynting factor which can be approximated by

$$\mathrm{PF}_i = \frac{V_i^{sat}\,(P - P_i^{sat})}{RT} \tag{5.4}$$

when assuming the molar volume of the liquid phase, V_i^{sat}, is independent of pressure.

At pressures above atmospheric, the vapor phase of systems such as these will involve deviations from the ideal gas law. The virial equation terminated at the second coefficient is a convenient and sufficiently accurate method for conditions up to a density of about one-half that of the critical. Suitable thermodynamic manipulation of this equation of state yields the vapor phase fugacity coefficient in terms of T, P, mole fraction, and pure and cross virial coefficients. Since the accuracy of the fugacity and compressibility are about the same for the pressure-explicit and density-explicit virial equations truncated at the second virial coefficient (Hayden and O'Connell, 1975) the most convenient form of the virial equation is:

$$z = \frac{PV}{RT} = 1 + \frac{BP}{RT} \tag{5.5}$$

V is the molar volume and, in a mixture of 2 components

$$B = \sum_{i=1}^{2} \sum_{j=1}^{2} y_i y_j B_{ij}(T) \tag{5.6}$$

B_{ij} is the second virial coefficient characterizing pair interactions between components i and j. The vapor fugacity coefficient is expressed

$$\ln \hat{\phi}_i = \left(2 \sum_{j=1}^{2} y_j B_{ij} - B \right) \frac{P}{RT} \tag{5.7}$$

$$= \left(B_{ii} + (1 - y_i)^2 \delta_{12} \right) \frac{P}{RT} \tag{5.8}$$

$$\delta_{12} = 2B_{12} - B_{11} - B_{22} \tag{5.9}$$

For the saturated vapor,

$$\ln \phi_i^{sat} = \frac{B_{ii} P_i^{sat}}{RT} \tag{5.10}$$

When all contributions are taken into account the fugacity factor becomes,

$$\Phi_i = \exp\left\{\frac{\left(B_{ii} - v_i^{sat}\right)\left(P - P_i^{sat}\right) + \left(1 - y_i\right)^2 \delta_{12} P}{RT}\right\} \tag{5.11}$$

however it may not always be possible or necessary. Typically ideal solution treatment is applied at elevated pressures ($\delta_{12} = 0$) and Poynting correction is assumed unnecessary ($PF_i = 1$).

$$\Phi_i = \frac{\phi_i^v}{\phi_i^{sat}} = \exp\left\{\frac{B_{ii}\left(P - P_i^{sat}\right)}{RT}\right\} \tag{5.12}$$

In the cases of real and ideal solution P and y_1 is calculated iteratively using:

$$P = \sum_{i=1}^{2} \frac{x_i \gamma_i P_i^{sat}}{\Phi_i} \tag{5.13}$$

$$y_1 = \frac{x_1 \gamma_1 P_1^{sat}}{\Phi_1 P} \tag{5.14}$$

When assuming ideal gas phase, ($\Phi_i = 1$), the pressure and vapor phase composition is determined by:

$$P = x_1 \gamma_1 P_1^{sat} + x_2 \gamma_2 P_2^{sat} \tag{5.15}$$

$$y_1 = \frac{x_1 \gamma_1 P_1^{sat}}{P} \tag{5.16}$$

The pure component second virial coefficients can be found in literature (Dymond and Smith, 1980) though access to data can be difficult for gasses or solvents not commonly used in the chemical industry. The search becomes almost impossible when looking for cross virial coefficients, B_{12}. There exists several methods for predictions of the second virial coefficient. One of the most accessible methods is the Pitzer Correlations where the critical temperature and pressure together with the acentric factor are needed (Smith *et al.*, 2005). Here we use the method of Hayden and O'Connell (1975) as a guide to realistic virial coefficients. The method only uses critical properties and molecular parameters, all of which may usually be estimated from molecular structure to the required accuracy.

5.2 Uncertainty Calculations

The uncertainty of the calculated pressures, σ_P, and vapor composition, σ_{y_1}, are calculated using the law of error propagation,

$$\sigma_P^2 = \sum_{j=1}^{n} \left(\frac{\partial P}{\partial a_j}\right)^2 \sigma_{a_j}^2 \tag{5.17}$$

$$\sigma_{y_1}^2 = \sum_{j=1}^{n} \left(\frac{\partial y_1}{\partial a_j}\right)^2 \sigma_{a_j}^2 \tag{5.18}$$

where a_j denotes any parameter applied in the calculations. The presented version of the rule of error propagation assumes the parameters are independent of each other which is not always the case for the modified Margules model. The covariance could be avoided completely by applying an orthogonal polynomial instead of the modified Margules model as G^E-model (Klaus and Van Ness, 1967). We have not done that here, so our uncertainties or error bars are approximations only.

From the modified Margules model 2, 4 or 5 terms are contributed where $\sigma^2_{a_j}$ has been estimated using the approach described in Section 4.2. If ideal or real solution treatment of the vapor phase is applied additional 2 or 3 contributions are included. The uncertainty of any second virial coefficient is difficult to estimate as experimental uncertainties are commonly not available (Dymond and Smith, 1980). The predictive method by Hayden and O'Connell (1975) gives average root-mean-square (RMS) deviations for different compound groups and also average RMS deviations for the cross second virial coefficient between different types of compound groups. The data from Hayden is used as rough estimates for the uncertainties however they have the weakness of being based on a small data set. The challenges are not problematic at the end because the derivatives in Equations 5.17 and 5.18 with respect to the second virial coefficients are generally small and their contributions are insignificant compared to those of the G^E-model.

The derivatives of P and y_1 with respect to parameters are extracted from the bubble point calculations, which is equivalent to solving:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \Delta f_1 \\ \Delta f_2 \end{bmatrix} = \begin{bmatrix} y_1 P \Phi_1 - P_1^{sat} \gamma_1 x_1 \\ y_2 P \Phi_2 - P_2^{sat} \gamma_2 x_2 \end{bmatrix} \tag{5.19}$$

with respect to P and y_1 with all parameters, $\underline{a}$, constant, whence

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} d\Delta f_1 \\ d\Delta f_2 \end{bmatrix} = \begin{bmatrix} \frac{\partial \Delta f_1}{\partial P} dP + \frac{\partial \Delta f_1}{\partial y_1} dy_1 + \sum_{j=1}^{n} \frac{\partial \Delta f_1}{\partial a_j} da_j \\ \frac{\partial \Delta f_2}{\partial P} dP + \frac{\partial \Delta f_2}{\partial y_1} dy_1 + \sum_{j=1}^{n} \frac{\partial \Delta f_2}{\partial a_j} da_j \end{bmatrix} \tag{5.20}$$

For each parameter j the partial parameter derivatives of P and y_1 can be determined from

$$\begin{bmatrix} \frac{\partial P}{\partial a_j} \\ \frac{\partial y_1}{\partial a_j} \end{bmatrix} = - \begin{bmatrix} \frac{\partial \Delta f_1}{\partial P} & \frac{\partial \Delta f_1}{\partial y_1} \\ \frac{\partial \Delta f_2}{\partial P} & \frac{\partial \Delta f_2}{\partial y_1} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial \Delta f_1}{\partial a_j} \\ \frac{\partial \Delta f_2}{\partial a_j} \end{bmatrix} \tag{5.21}$$

$\partial \Delta f_i / \partial a_j$ are derived from Equation 5.19 and the expressions used for calculation of Φ_i and γ_i. The method has been applied in all case studies to calculate error bars for xP- and xy-plots.

6

Optimization of Parameters for Force Field Descriptions

The problem of missing parameters is frequently encountered in mathematical modeling of chemical systems. For G^E-models like UNIFAC an extensive volume of binary experimental data is required for extensions of the parameter matrix. The situation is less problematic in the area of molecular modeling. The force field parameters can be optimized by reproducing a combination of *ab initio* calculations and pure component properties which makes it possible to extend simulations to new compounds and compound groups without the need of binary mixture data. Apart from a minimum of experimental data, a computational chemistry program and simulation time are needed.

In section 2.1 it was briefly described how Siepmann *et al.* and Ungerer *et al.* determined the parameters for their respective force fields. One could say that there exists as many approaches to optimization of parameters as the number of force fields, but some guide lines exists for determination of new parameters. Three examples of force field potentials have been presented already. These are the TraPPE-UA potential (Eq. 2.4 on pages 8), the AUA potential (Eq. 2.5 on pages 9), and the CHARMM potential (Eq. 3.13 on pages 20). The force field potential can roughly be resolved into parts describing inter- and intramolecular phenomena. A general overview is given of the different approaches for determining both parts below and finally the simple parameter optimization approach applied in this work is described.

6.1 Determination of Intramolecular Potential

The intramolecular potential is a simplified descriptions of intramolecular forces within a single molecule. In most force fields quantum mechanical *ab initio* calculations are reproduced in the optimization of the force field. In the quantum mechanical calculations the equilibrium structure and energy of the molecule are determined using the Schödinger equation (McQuarrie, 2000). This is followed by a series of calculations where a specific characteristic feature is modified and constrained and the structure with the lowest energy is determined again. A classical example is n-butane where conformational energy of the C-C-C-C dihedral angle is determined by constraining or fixing the angle and

calculating the energy of the molecule (for example Smith and Jaffe (1996)).

The output combinations of coordinates and energies are then to be reproduced by the force field potential. As shown in the examples listed above there are many possible contributions and mathematical expressions for each contribution. The basics are the bond angle and dihedral angle terms. The harmonic bond stretch potential is typically left out of the force fields developed for Monte Carlo simulations where united atom types descriptions are applied. The CHARMM force field is special by also including the 1-3 Urey-Bradley term (A. D. MacKerell *et al.*, 1998). An indirect contribution to the intramolecular potential is the van der Waals and electrostatic forces within the molecule which are normally considered intermolecular forces. Normally the potential applied for the non-bonded van der Waals and electrostatic forces is used for 1-4 interactions and with atomic centers at further distance within the molecule.

The intramolecular potential does not play an important role for small rigid molecules, but for larger molecules the distribution of the intramolecular configurations can be important. This is seen when dynamic properties are studied, for example in the case of proteins.

The newer force fields, like TraPPE-UA and AUA, have reused previous determined parameters for the intramolecular potential by other groups. TraPPE-UA borrowed from OPLS (Briggs *et al.*, 1991) and the AUA force field of Ungerer *et al.* (2000) borrowed from previous AUA work of Toxvaerd (1997).

6.2 Determination of Intermolecular Potential

The optimization of the intermolecular part of the potential is where the difference between the force fields mentioned at the start of the chapter becomes clear. The ambitions of the force field developers have a clear connection with the evolution of computers. In the earliest work, Jorgensen (1981a,b,c) started by reproducing the interactions of dimers and then testing the parameter set for few examples of liquid phase simulation because computers were limited and slow compared to modern standards. Later the OPLS-UA force field was developed where liquids were simulated to reproduce molar volume, V_m, and enthalpy of vaporization, ΔH_{vap}, (Jorgensen *et al.*, 1984). Ten years later the PC has made computer power more accessible and Toxvaerd (1990) was able to make simulations for several compounds at a wide range of state points. Toxvaerd (1990, 1997) was interested in the simulation of liquid alkanes from an "Equation of State " point of view where volume and temperature were fixed and the pressure was predicted. Thereby a larger phase space area was covered with respect to temperature and pressure compared to earlier work. With the development of the Gibbs Ensemble Monte Carlo methods (see Section 2.1) it became possible to include phase equilibrium properties in the objectives (Martin and Siepmann, 1998; Ungerer *et al.*, 2000; Errington and Panagiotopoulos, 1999a). This latest development is mostly due to developments in theory and

novel algorithms and not based on computer developments.

The list of examples is already long and presents the chronological development of one main branch of the force fields. The other main branch is interested in biological phenomenons where proteins and cell membranes are simulated in a water or vacuum environment (Jorgensen and Swenson, 1985; A. D. MacKerell *et al.*, 1998; Foloppe and Mackerell, 2000). Opposite to the work above only ambient temperatures and pressure are of interest however in the biology field dynamic properties plays an important role.

6.2.1 Transferability of parameters

There are two main aspects of the phenomenon. In the first case parameters are for example optimized for alkanes using propane, pentane and octane. If the determined parameter set is transferable, it can be used to predict behavior of butane or decane. The second example could be alkenes where the saturated part of the molecule is described using previously determined alkane parameters and new parameters are only optimized for the double bonded carbon atoms. This sequential approach have been applied most common force fields like OPLS (Jorgensen *et al.*, 1984; Jorgensen, 1986), TraPPE-UA (Martin and Siepmann, 1998, 1999; Chen *et al.*, 2001), and AUA (Ungerer *et al.*, 2000; Kranias *et al.*, 2003). The approach also helps to reduce the large parameter space available when comparing the number of parameters possible against the experimental and *ab initio* data available.

6.2.2 Partial Charges Determination

There are generally three approaches for determining the atomic charges which are all based on quantum chemical calculations using software like Gaussian or Spartan. In early work by Jorgensen (1981a,b,c) the initial set of van der Waals parameters and charges where determined by reproducing dimer energies from quantum chemical calculations at different distances and angles. The charges generated has later been used in the OPLS-UA (Jorgensen, 1986) and TraPPE-UA (Chen *et al.*, 2001) force fields.

In the CHARMM force field the atomic charges for the electrostatic charges are determined by reproduction of the aqueous solvation energies (A. D. MacKerell *et al.*, 1998).

The last approach stands out, because quantum mechanical calculations is reduced to calculation of the electrostatic potential around the molecule called restrained electrostatic potential (RESP) method (Bayly *et al.*, 1993; Levy and Enescu, 1998). The electrostatic potential is calculated at points in shells at growing distance to the molecule. The atomic or partial charges are then calculated by the optimal reproduction of the electrostatic potential of all the points. The calculation can be expanded to an average of the charges determined for the molecule in different local equilibrium structures together with the global equilibrium structure as the different structures give varying results.

6.2.3 Time Consumption

The time consumption or cpu-time of a general optimization is in most cases longer and two factors count for the majority of the time used. These are the amount of data used in the evaluation of the objective function and the time for prediction of a single point by the model. In a simple optimization the calculations take less than one second, for example the optimization of G^E-parameters described in Chapter 4. In the G^E-model UNIFAC thousands of data points are included in the optimization and even though the prediction of each data point is fast, the time consumption is counted in days rather than seconds. The long prediction time of each data point is typically seen in dynamic models, for example determination of kinetic parameters for a model of a chemical reaction, where the evaluation at each iteration includes integration with some partial differential equation solver.

The optimization of parameters for a force field is an extreme example of the latter case because of the simulation time. There are several options of the optimization which can make the evaluation time at each iteration range from days to weeks. These are the number of compounds, state points (T,P) and the number of properties (V_m ΔH_{vap}, P_{vap}...). The ability to simulate a single compound or a whole compound group makes a big difference which depends on the specification set in the objective of the force field. Liu *et al.* (2004) only simulated 2 compounds with individual parameters sets where the MD simulation include 2 temperatures for each compound to reproduce enthalpy of vaporization and density. A larger effort has been used in the n-alkane force field description by Ungerer *et al.* (2000) where vapor pressures, vaporization energies and liquid densities of ethane, n-pentane and n-dodecane at several temperatures have been included in the objective function. The optimization of the 6 parameters took 1200h or 50 days (Ungerer *et al.*, 2000) which has to be considered a short time as good start estimates were available from the work of Toxvaerd (1997). Another type of time influencing option is the number of molecules included in the simulations because there is a clear connection between the size of the system and the noise of the predicted properties. When calculating the derivative of the objective function which is applied in the determination of the changes in the parameters for the next iteration the noise or uncertainty of the predicted properties can result in a wrong sign of the derivative thereby making the optimization go in the wrong direction. On the other hand simulation of too big systems will result in long simulation time which will delay the final parameter set. Consequently it is important to choose a system size where the noise or uncertainty is low enough for being applicable in the optimization.

6.3 Simple Approach

In the optimization of force field parameters of this work a simple approach has been applied. The overall objective of the approach is to apply the ca-

pabilities available to an engineer who does not have endless resources i form of experience, special software and time. The basis available for the engineer includes publicly available parameter tables and software which can be downloaded without cost.

Several shortcuts have to be used when specialty software is not applied in the determination of intramolecular parameters and partial charges. Consequently, equivalent intramolecular parameters are applied when some are missing and partial charges are taken from similar structures. The effort is put into optimization the Lennard-Jones parameters of the intermolecular potential.

When simplifying the procedure for determination of the parameters assumptions are made which are not completely bulletproof. The consequence can be a force field description unable to predict the behavior of the compound or compound group with respect to temperature dependency or only density can be predicted but not enthalpy of vaporyzation. The approach has been applied in two case studies, (2H)-heptafluoropropane and lactone compound group, which can be found in Chapter 7.

6.3.1 Simulations

Liquid systems containing 216 molecules were used which is equivalent to the work of Chen *et al.* (2002); Liu *et al.* (2004). Isobaric-isothermal ensemble was applied in MD simulations with initial minimization the system followed by equilibration for 1 ns and a production period is 5 ns with a time step of 1 fs. The total energy and simulation box size were sampled each 500 fs.

6.3.2 Optimization Procedure

The objective function includes densities at selected temperatures and in some cases enthalpy of vaporization, ΔH_{vap}, at 298.15 K.

$$SS = \sum \left(\frac{\rho_{exp} - \rho_{MD}}{\rho_{exp}} \right)^2 + \sum \left(\frac{\Delta H_{vap}^{exp} - \Delta H_{vap}^{MD}}{\Delta H_{vap}^{exp}} \right)^2 \tag{6.1}$$

ΔH_{vap} is calculated in a simplified fashion where the intermolecular interaction energy of the vapor phase is considered negligible together with the difference in the intramolecular energy of the liquid and vapor phases. These approximations are acceptable at ambient conditions. The expression is:

$$\begin{aligned} E_{inter}^{vap} &\approx 0 \\ E_{intra}^{vap} &\approx E_{intra}^{liq} \\ \Delta H_{vap}^{MD} &\cong -E_{inter}^{liq} + RT \end{aligned} \tag{6.2}$$

The average and variance of the density was calculated from the simulation box side length sampled during the production period.

$$\rho_{avg} = \frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} \rho_i \tag{6.3}$$

$$(\rho^2)_{avg} = \frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}} \rho_i^2 \tag{6.4}$$

$$\sigma_\rho^2 = (\rho^2)_{avg} - (\rho_{avg})^2 \tag{6.5}$$

The optimization procedure starts with a set of start estimates and the partial derivatives of the objective function is determined by performing a series of simulations where a single parameter, a_j, is modified in each simulation and the rest of the parameters are kept constant. The numerical derivatives at iteration I are calculated by:

$$\frac{dSS^I}{da} = \sum_{n_{points}} \frac{SS\left(a_j^I + \delta\right)_{a_{h\neq j}^I} - SS^I}{\delta}; \quad j = 1 \ldots n_{par} \tag{6.6}$$

The derivatives are then applied in the calculation of the parameter set of the next iteration, $I + 1$. The number of simulations needed at each iteration is a product between the number of points in the objective function and number of parameters, $n_{points} \cdot (n_{par} + 1)$.

7

Case Studies

The presented case studies can be divided into small groups. The two first cases, benzene - methyl acetate and methyl acetate - acetone, show close to ideal behavior. The rest of the systems studied are of a more non-ideal class, starting with benzene - ethanol where some self association occurs between the ethanol molecules and continuing with methyl acetate - n-pentane. The fifth system originates from the third Industrial Fluid Phase Simulation Challenge where a mixture of ethanol and 1,1,1,2,3,3,3-heptafluoropropane is studied. Experimental bubble point pressure data was given at a low temperature and a high temperature bubble point pressure curve had to be predicted. The case study is a good example of handling large temperature scales. The last case study covers a whole compound group, lactones, which are cyclic esters. The objective is to study how the force field parameter transferability concept can be explored together with FST analysis when studying a novel compound group.

CHARMM parameter tables applied in all simulations are listed in Appendix C and correlation parameters for reverse approach calculations are in Appendix D.

7.1 Benzene - Methyl Acetate

The system was selected for the initial case study because it was considered not to contain any special phenomena. The system is also a good test because it allows the study of the behavior of benzene with a linear ester, as mixtures of lactones with benzene is going to be presented later. From the experimental data it is known that the system shows close to ideal behavior. The temperature is 303.15 K and the experimental pressures are all below atmospheric pressure, which allows the assumption of ideal gas behavior of the vapor phase. The CHARMM force field descriptions of the compounds are taken from original work of A. D. MacKerell *et al.* (1998). The atomic descriptions are shown in Figure 7.1. MD simulations were performed for 7 compositions covering the bulk composition range. 1000 molecules were included in the simulation box and standard settings from section 3.2 were applied.

Pure component simulations were performed to predict the molar volumes, V_m^{MD}, to evaluate the CHARMM force field description of components. The

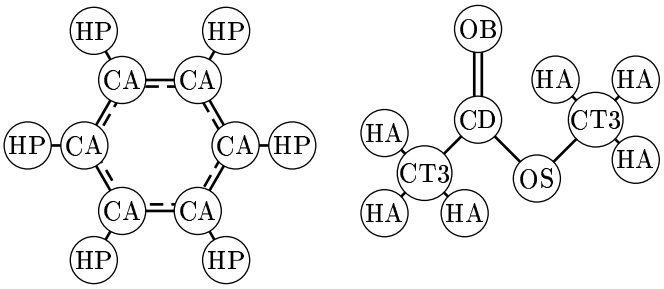


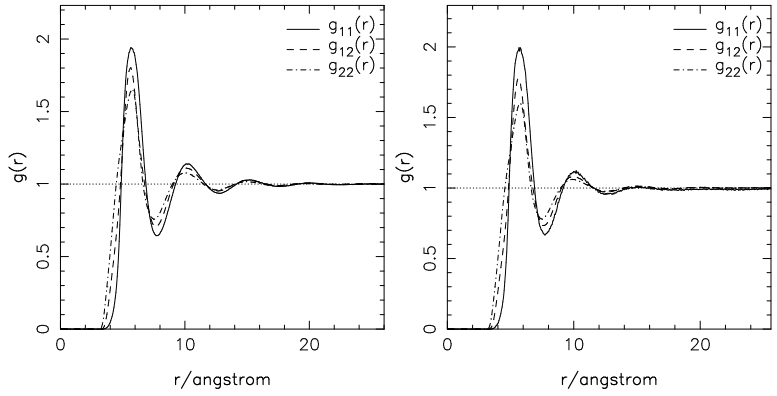
Figure 7.1. Atomistic CHARMM description of benzene and methyl acetate

compound	$V_m^{exp}/\mathrm{cm}^3/\mathrm{mol}$	$V_m^{MD}/\mathrm{cm}^3/\mathrm{mol}$	rel. error %
benzene	89.961	92.82±0.54	3.2±0.6
methyl acetate	80.442	81.75±0.45	1.6±0.5

Table 7.1. Comparing pure compound molar volumes for benzene and methyl acetate at 303.15 K.

results are listed in Table 7.1 and V_m^{MD} of methyl acetate is within the 2% considered acceptable. The molar volume of benzene is slighty too high, however the relative error is not considered critical.

The radial distribution functions show a simple characteristic behavior where all of the direct interaction peaks (first maximum) are made up of an average of several intermolecular angles but neither 11,12 nor 22 seems to dominate (see Figure 7.2). The benzene-benzene radial distribution function ($g_{11}(r)$) is independent of the composition which can be seen when comparing solid lines of Figures 7.2.a and 7.2.b where both extrema have more or less the same values. The behavior is clearly shown in Figure 7.3 where the TCFIs of benzene (H_{11}) have the value -0.9 across the composition range. Methyl acetate show a small tendency to cluster when becoming diluted. In figure 7.2 the maximum of the methyl acetate - methyl acetate rdf (g_{22}^{max}) changes from 1.65 at equimolar conditions to 1.6 at $x_B = 0.125$ and H_{22} goes from -0.95 to a value in the vicinity of zero. The TCFIs of benzene - methyl acetate, H_{12}, show an average behavior between H_{11} and H_{22}, although dominated by the benzene - benzene behavior. The difference between the TCFIs, ΔH, are all positive but dominated by large relative uncertainties. The results of the reverse approach are reproduced extremely well for methyl acetate rich compositions, but the error increases fast as methyl acetate becomes dilute. This can be explained by the large error in H_{22} at high benzene concentrations. There are also errors in H_{11} and H_{12} in the dilute benzene region, however they cancel out each other. The TCFIs displayed in Figure 7.3 have been



(a) $x_1 = 0.5$ (b) $x_1 = 0.125$

Figure 7.2. Radial distribution functions for benzene (1) - methyl acetate (2) system at T = 303.15 K and 1000 molecules included. solid line: benzene-benzene (11), dashed line: benzene-methyl acetate (12) and dot dashed line: methyl acetate-methyl acetate (22).

generated using method 3 from section 3.3 where the TCFIs are a sum of three contributions (Eq. 3.22). In Figure 7.4 a comparison is made between method 1 and 3. In the given case the superiority of method 3 is obvious. The uncertainties introduced by the truncation of method 1 are big relative to the order of magnitude of the TCFIs which gives especially bad results for dilute benzene compositions. The derivative of the activity coefficient from both method 1 and 3 are used to determine modified Margules parameters which are applied to predict the activity coefficient curves ($\ln \gamma(x)$) in Figures 7.4.c and 7.4.d. The mM γ-prediction of method 3 follows the γ-behavior of the experimental data with a minor deviation for γ_2 in the methyl acetate dilute region. The mM γ-prediction of method 1 are good for benzene except in the dilute part where the predicted activity coefficients are much larger than the experimental values. This is a direct consequence of the significant error in $d \ln \gamma_1 / dx_1$ at $x_1 = 0.125$. The γ-prediction of method 1 for methyl acetate are larger than the experimental data through the whole composition range. For

	Method 1	Method 3
A_{12}	0.691±0.082	0.262±0.027
A_{21}	0.06±0.18	0.361±0.060

Table 7.2. Modified Margules parameters for benzene (1) - methyl acetate (2) mixture at 303.15 K determined from FST analysis using method 1 and 3 for integration of rdfs.

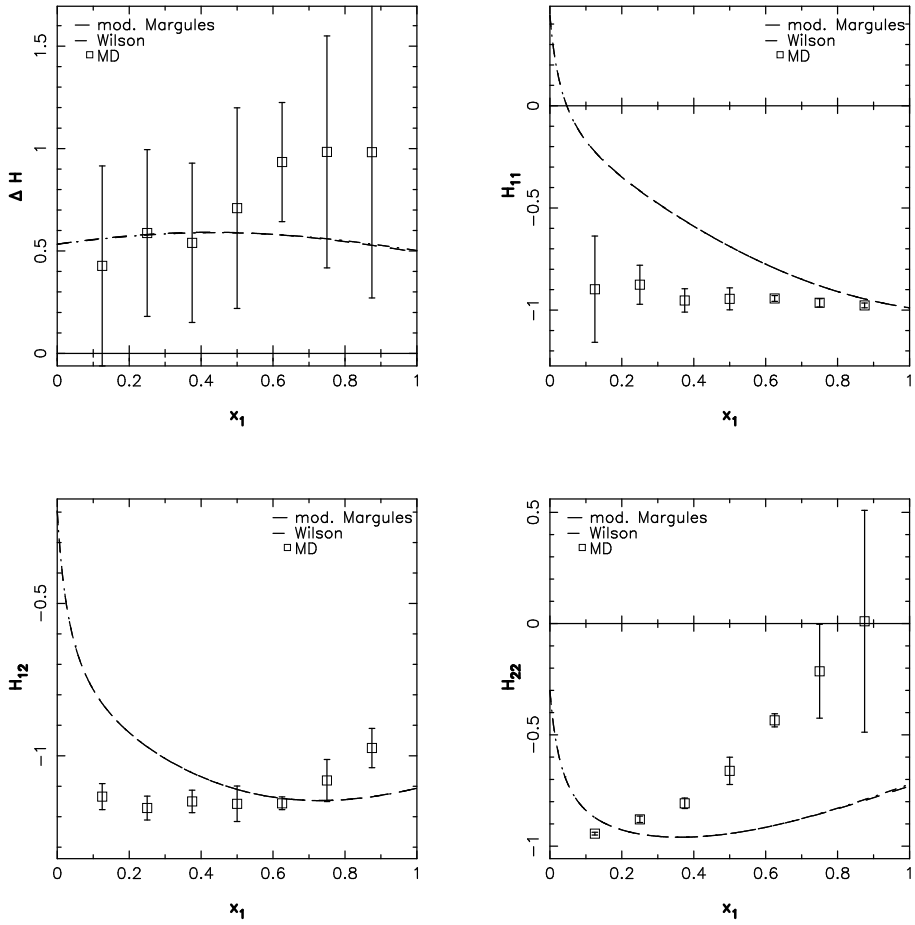


Figure 7.3. Total Correlation Function Integrals (TCFIs) for benzene (1) - methyl acetate (2) system at 303.15 K. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dot-dashed: Wilson.

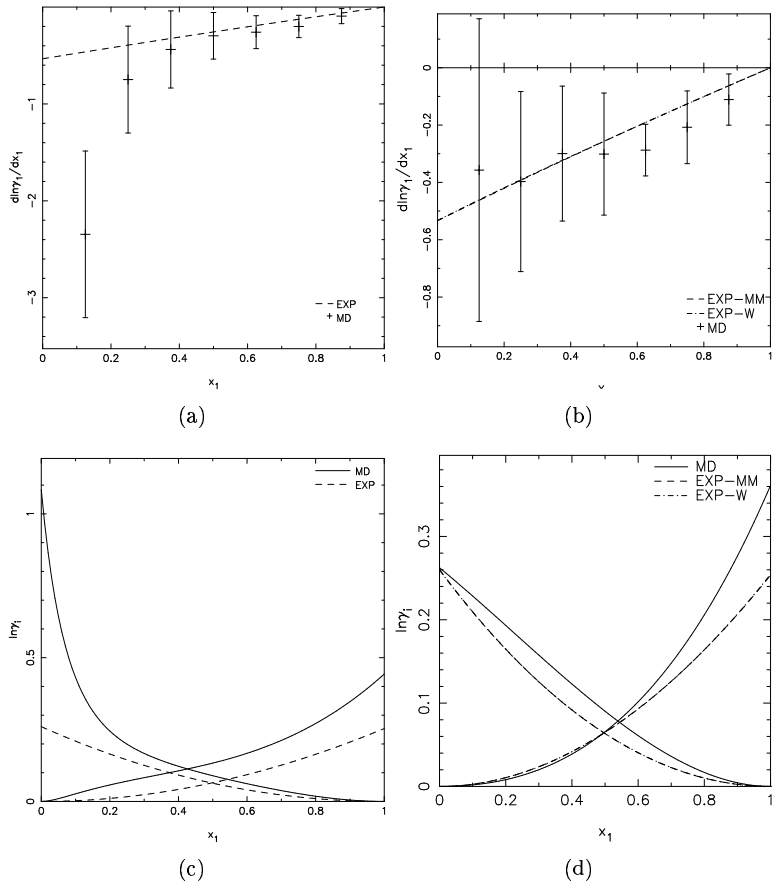


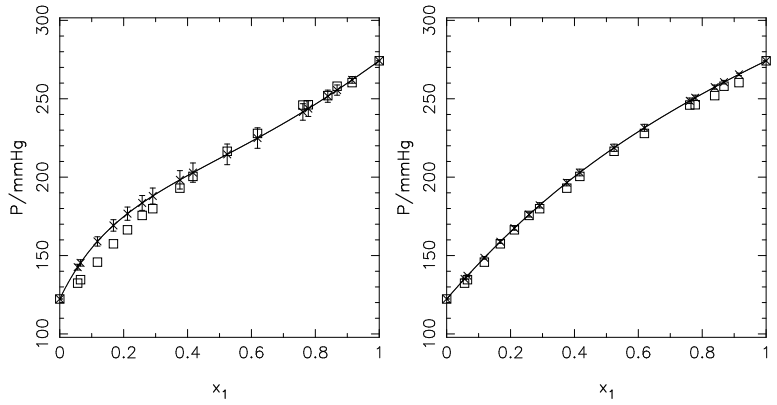
Figure 7.4. Benzene(1) - methyl acetate(2) at $303.15K$. Comparing results of MD results with modified Margules correlation. Figures a and c use the TCFIs found using method 1. Figures b and d use the TCFIs found using method 3. Crosses and solid lines are results from MD simulations. Dashed lines are generated by modified Margules where the parameters have determined using experimental data.

both method 1 and 3 the simplest form of the modified Margules expression gave the optimal parameter set. In the case of method 1 a significant reduction was made in the value of the objective function when expanding the modified Margules expression to 4 parameters, however the uncertainties of 3 of the 4 parameters were large in magnitude than the parameters themselves, and the parameter set was discarded. The modified Margules parameter sets are listed in Table 7.2.

The modified Margules parameter sets from Table 7.2 are combined with the assumption of ideal gas behavior of the vapor phase and the bubble point pressure curves are predicted (see Figure 7.5). The scores were calculated according to Equation 5.1 using the data of Nagata *et al.* (1973). The results are listed in Table 7.3 together with the score if ideal liquid phase is assumed ($\gamma_i(x) = 1$). The score when applying the method 3 mM parameter set is very

	method 1	method 3	ideal liquid
SCORE	3.63	1.41	4.39

Table 7.3. SCORES (Eq. 5.1) for benzene - methyl acetate system at 303.15K where method 1 and 3 had been used for the integration of rdfs in the FST analysis and for reference the score if the liquid phase is assumed to be ideal.



(a) method 1 (b) method 3

Figure 7.5. Bubble point pressure curves for benzene (1) - methyl acetate (2) system at 303.15K using 2 sets of modified Margules parameters where method 1 and method 3 has been applied for integration of rdfs in FST analysis.

good and is better than one could expect for a system showing close to ideal behavior. The errors and uncertainties of H_{ij}'s seem to cancel out each other instead of increase the error of ΔH. The relative large error of ΔH at $x_1 > 0.5$ is damped when converted to derivatives, $d \ln \gamma_1 / dx_1$, and consequently the

modified Margules description becomes good and the predictions and score are very satisfying. The relative large uncertainties of the TCFIs, $\sigma_{H_{ij}}$, in Figure 7.3 indicate the result could have been significantly different and consequently less precise.

The score of the method 1 integration approach might seem good however when compared with the score of pressure prediction when assuming ideal liquid phase the results are not good. The score when assuming ideal liquid phase is acceptable from a design point of view and the small ΔH's along the whole composition range can be used as an indication that one can assume ideal liquid phase and get pressure predictions which are acceptable.

7.2 Methyl acetate - acetone

The methyl acetate - acetone system is a difficult challenge from the FST analysis point of view. The system shows close to ideal behavior, but includes an azeotropic point. The temperature is elevated to 323.15K compared to previous systems but experimental pressures are still below atmospheric pressure and the gas phase is considered ideal. The methyl acetate description is the same as in the previous case study. The description of acetone is taken from work of Martin and Biddy (2005) where they used it for predictions of heats of vaporization. The atomic CHARMM descriptions are shown in Figure 7.6. MD

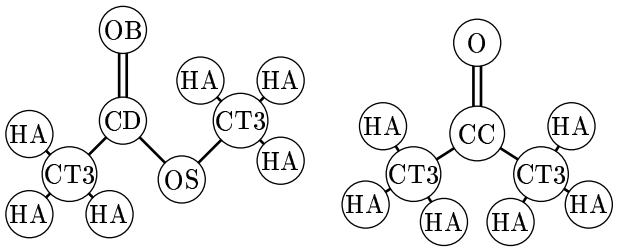


Figure 7.6. Atomistic CHARMM description of methyl acetate and acetone

simulations were performed for 9 compositions across the composition range. 512 molecules used for the mixtures and standard settings from section 3.2 were applied.

Pure component simulations were performed to predict the molar volumes, V_m^{MD}, to evaluate the CHARMM force field description of components. The predicted molar volumes (Table 7.4) are within the 2%-limit for acetone while V_m of methyl acetate is slighty too high. But the inaccuracy is not critical.

Figure 7.7 illustrates the large variance of the like-like TCFIs, $\sigma^2_{H_{ii}}$, in the dilute composition area for the methyl acetate - acetone system. H_{11} and H_{22}

compound	$V_m^{exp}/\mathrm{cm}^3/\mathrm{mol}$	$V_m^{MD}/\mathrm{cm}^3/\mathrm{mol}$	rel. error %
methyl acetate	82.9	84.9±1.3	2.4±1.5
acetone	76.66	77.9±1.3	1.6±1.7

Table 7.4. Comparing pure compound molar volumes for methyl acetate and acetone at 323.15 K.

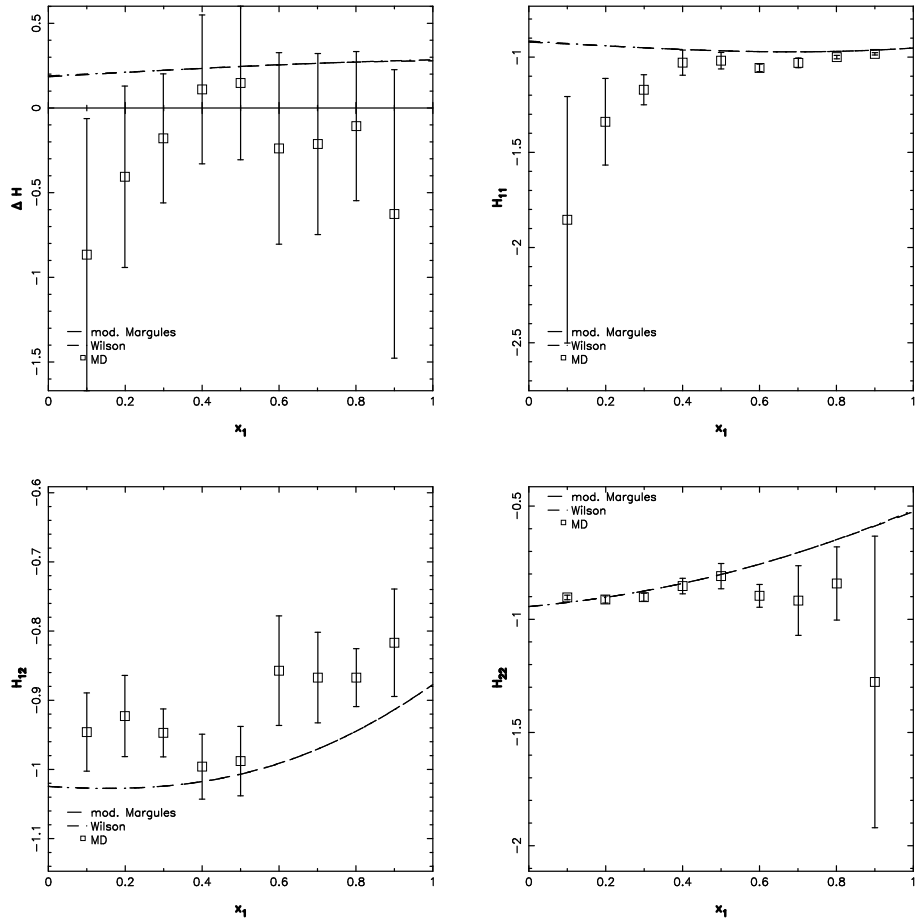


Figure 7.7. Total Correlation Function Integrals (TCFIs) for methyl acetate (1) - acetone (2) system at 323.15 K using integration method 3. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dashed-dotted: Wilson for G^E-model.

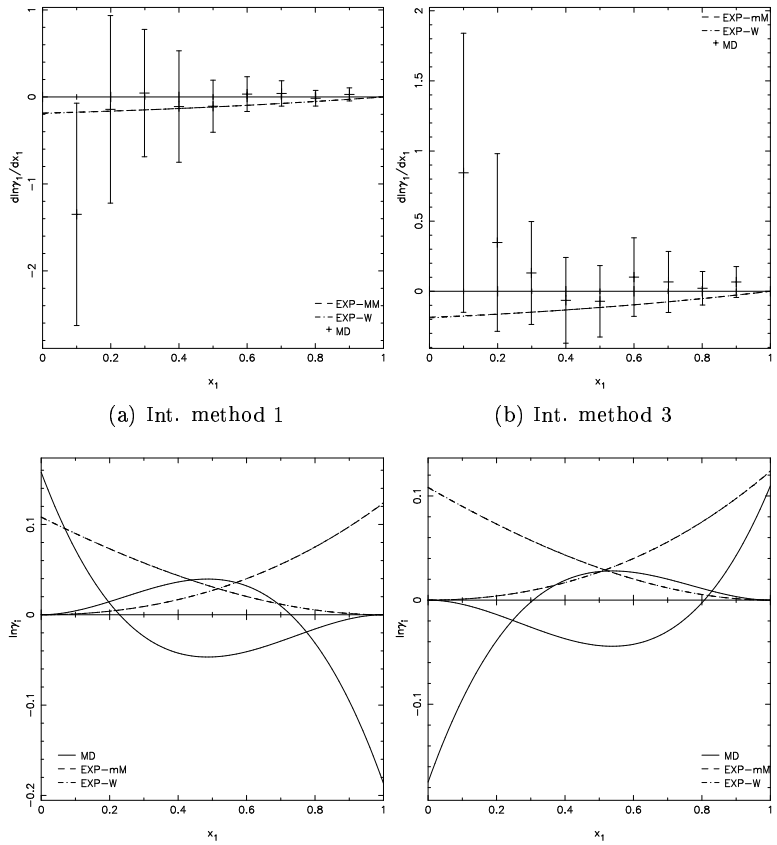
have relative large error bars at low concentrations. $\sigma^2_{H_{12}}$ shows less composition dependence. The behavior resembles the results of the benzene - methyl acetate system (Figure 7.3). These effects are clearly shown in the ΔH-plot where the variances, $\sigma^2_{\Delta H}$ at $x_1 = 0.1$ and $x_1 = 0.9$ are significantly larger. The dashed and dashed-dotted lines show the TCFIs generated by the reverse approach of Wooley and O'Connell (1991) using the Wilson and modified Margules models. The TCFIs from MD simulation generally reproduce the reverse approach results however the like-like integrals, H_{ii}, are badly reproduced in their dilute area and they have large variances. The ΔH-values are generally small indicating an almost ideal system. The deviations from reverse approach results at dilute conditions are small on an absolute scale. ΔH goes from negative to positive and back to negative again and this fluctuating behavior gives muddy picture together with the relatively large standard deviation of ΔH. If the standard deviations are taken to the extreme the ΔH-curve could have been all negative or almost all positive. This is the best example of the weakness the methodology when treating systems showing close to ideal behavior, however one still gets acceptable results by assuming ideal behavior of the liquid phase as the mixture behaves nearly ideal and the pure component vapor pressures are almost uniform.

The fluctuations of ΔH around zero results in similar behavior of $d\ln\gamma_1/dx_1$ for the integration method 3 results in Figure 7.8.b. At dilute methyl acetate conditions the predicted $d\ln\gamma_1/dx_1$-values are positive while the behavior of $d\ln\gamma_1/dx_1$ extracted from the experimental data (Olson, 1981) gives small negative values. The disagreement is visualized in Figure 7.8.d where the $\ln\gamma_1^{MD}$-curve has a small negative value at $x_1 = 0.1$ while the experimental activity coefficient has a small positive value. The results of integration method 1 is for most of the composition range similar to the results of method 3 with fluctuations around zero. The significant difference is at $x_1 = 0.1$ where $d\ln\gamma_1/dx_1$ has a relatively large negative value (Fig. 7.8.a) and a negative value at $x_1 = 0.8$. The $\ln\gamma_i^{MD}$-curves of integration method 1 (Fig. 7.8.d) fluctuates differently than method 3 results but all results are in the vicinity of zero.

The relative large fluctuations in Figure 7.8.b suggests the alternative objective function, SS_2, (Eq. 4.3) could be preferable. The resulting parameters are listed in Table 7.5 together with the optimized parameter sets of SS_1 (Eq. 4.2). In all three cases the increase from 2 to 4 parameters significantly reduced

	Method 1	Method 3	
	SS1	SS1	SS2
A_{12}	0.158±0.099	-0.175±0.063	0.037±0.082
A_{21}	-0.19±0.22	0.11±0.14	-0.077±0.039

Table 7.5. Modified Margules parameters for methyl acetate (1) - acetone (2) mixture at 323.15 K determined from FST analysis using objective functions SS1 (Eq. 4.2) and SS2 (Eq. 4.3) for results of integration method 1 and 3.



(a) Int. method 1 (b) Int. method 3

(c) Int. method 1 (d) Int. method 3

Figure 7.8. $d\ln\gamma_1/dx_1$ and $\ln\gamma_i$ as function of x_1 for methyl acetate (1) - acetone (2) at 323.15 K. Solid (modified Margules) and dashed (Wilson) are determined from experimental data (Olson, 1981). (a) and (c) are results of applying method 1 integration of rdfs and (b) and (d) are results for integration method 3.

the objective function value, but in all cases the uncertainty or standard deviation of the parameters were too large for the 4-parameter description. For 5-parameter descriptions the uncertainties were again larger than the parameters themselves and the simplest form of the modified Margules expression was selected.

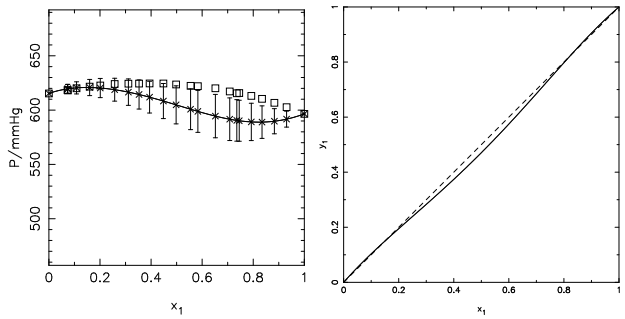
The bubble point pressure curves in Figure 7.9 might seem suspicious because both a maximum above and a minimum below the ideal mixture pressure line exists. The phenomenon observed for the results of the FST analysis is called double or poly azeotropic behavior which can occur for mixtures close to ideal conditions. The first experimental observations of the phenomenon were made by Duncan *et al.* (1968) which is quite late in the history of experimental chemistry. The system was benzene with hexafluorobenzene and poly azeotropes have since been observed by other workers (Leu and Robinson, 1991; Christensen and Olson, 1992; Burguet *et al.*, 1996; Aucejo *et al.*, 1996, 1997; Kao *et al.*, 1997). The outcome of applying the alternative objective function, SS_2, did neither give significantly better nor did it return unreliable results. The resulting scores when applying SS_1 and SS_2 are almost the same and a double azeotrope is predicted in both cases. The application of SS_2 is in general advisable when the standard deviation of the TCFIs, $\sigma^2_{\Delta H}$, are relatively large compared to the TCFIs themselves. SS_2 should not be applied instead of, but as a supplement to SS_1. Then the resulting predictions can be compared and disagreements should lead to additional simulation or analysis.

Integration method	objective function	SCORE	$x_{\text{azeotrope}}$ 1	2
1	SS1	2.29	0.149	0.823
3	SS1	2.54	0.318	0.758
	SS2	2.36	0.026[a]	0.863
Exp.		$3.4570 \cdot 10^{-6}$	0.380	

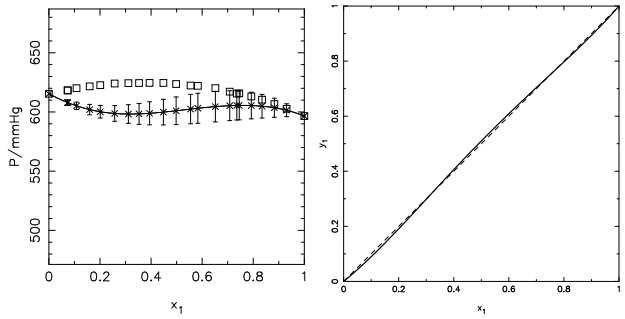
Table 7.6. SCORES (Eq. 5.1) and azeotropic points for methyl acetate - acetone system at 323.15K applying parameter sets of Table 7.5. [a] From $x_1 = 0.0$ to $x_1 = 0.026$ the compositions of liquid and vapor phase are almost identical, $x_1 \approx y_1$.

As mentioned in the introduction to the methyl acetate - acetone system at 323.15 K the experimental data set only includes a single azeotropic point which is listed in Table 7.6 together with the predicted double azeotropes.

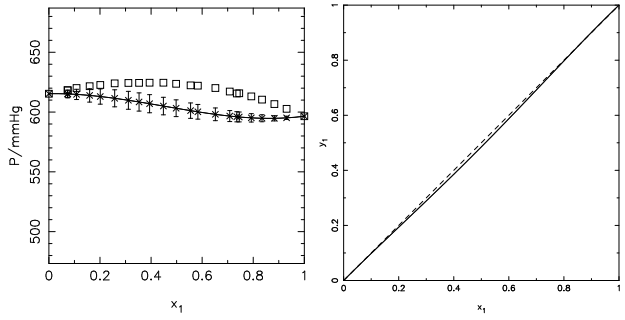
From a negative point of view one could conclude, the single azeotrope was not predicted by the FST analysis approach. However all three approaches within the FST analysis methodology predicted a close to ideal system with scores not higher than 2.54. Azeotropic behavior was detected although more azeotropic points were found from simulations than from experiments.



(a) SS_1 & method 1 (b) SS_1 & method 1



(c) SS_1 & method 3 (d) SS_1 & method 3



(e) SS_2 & method 3 (f) SS_2 & method 3

Figure 7.9. Predicted xP and xy diagrams using modified Magules parameter set from Table 7.5 for methyl acetate (1) - acetone (2) at 323.15 K.

7.3 Benzene - Ethanol

The benzene - ethanol system is moderately non-ideal and is studied at 298.15 K where the vapor phase can be considered an ideal gas. The CHARMM descriptions were taken from literature (A. D. MacKerell *et al.*, 1998) and the atomic descriptions are shown in Figure 7.10. MD simulations were performed

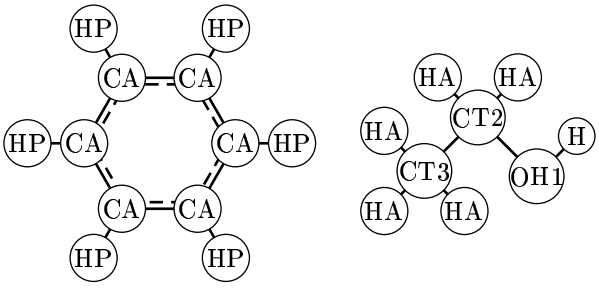


Figure 7.10. Atomistic CHARMM description of benzene and ethanol.

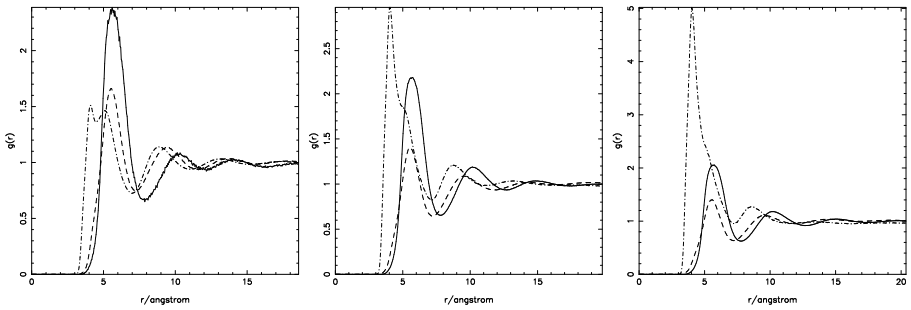
for 9 compositions across the composition range. 512 molecules used for the mixtures and standard settings from Section 3.2 were applied in simulations. Pure component simulations were made with 216 molecules to predict molar

compound	V_m^{exp}/cm^3/mol	V_m^{MD}/cm^3/mol	rel. error %
benzene	89.502	92.08±0.72	2.9±0.8
ethanol	58.516	59.56±0.57	1.8±1.0

Table 7.7. Comparing pure compound molar volumes for benzene and ethanol at 298.15 K.

volumes (Table 7.7). V_m^{MD} of ethanol is within the acceptable error of 2% which is not seen for benzene, however the level of inaccuracy is consistent with the result at 303.15 K from Table 7.1.

In Figure 7.11 three sets of radial distribution functions are displayed where the change in the rdf of ethanol, $g_{22}(r)$, across the composition range is dramatic. The direct interaction part of $g_{22}(r)$ is characterized by double peak where the first peak is a dimer from self association and the second is the average of the rest of the interaction angles. At high concentration of ethanol the peaks are equally strong, but as the mole fraction of benzene increase the self association becomes dominant but the maximum of both peaks increases and the distributions become narrow. The effect of this behavior is captured in H_{22} in Figure 7.12 where H_{22} goes to extremely high values at small ethanol concentrations. There is good agreement between the H_{22}'s from MD simulations



(a) $x_1 = 0.1$ (b) $x_1 = 0.5$ (c) $x_1 = 0.7$

Figure 7.11. Radial distribution functions for benzene (1) - ethanol (2) system at T = 298.15 K and 512 molecules included. solid line: $g_{11}(r)$, dashed line: $g_{12}(r)$and dot dashed line: $g_{22}(r)$.

and those predicted from the reverse approach. The high absolute uncertainty of H_{22} at $x_1 = 0.9$ becomes less significant because the relative uncertainty is low. Even though the absolute uncertainty is high it is within the boundaries of the disagreement between the results of the Wilson and modified Margules models in the reverse approach results. H_{22} in the benzene - ethanol case is quite special for reverse approach results because one rarely can observe such a large difference between the position and value of the maximum of a TCFI. The RDFs of benzene-benzene, $g_{11}(r)$, are less sensitive to composition. The change can be seen in Figure 7.12 where H_{11} small positive values for most of the composition range to negative values at high concentration of benzene. As in the case of ethanol, H_{11} from MD simulations follow the trend of the reverse approach results. At dilute benzene compositions the MD results are too small for H_{11} however ΔH is dominated by terms from H_{12} and H_{22} and there it has no effect on the final result. The behavior of H_{12} is reproduced quite well for most of the composition range, though the minimum of the reverse approach around $x_1 = 0.65 - 0.70$ is not predicted by the MD simulations. The consequence is seen in ΔH in the composition range $x_1 = 0.6 - 0.80$ where ΔH from the MD simulations is underestimating the reverse approach results. The general agreement between ΔH of simulations and reverse approach is good.

System size effects were studied as the equivalence of ensembles (NPT, μVT) is only true for infinitely large systems. It is also true for smaller systems, but the fluctuations are larger for the smaller systems, and the systems may be so small that a thermodynamic description is meaningless. We are not at that limit here. Table 7.8 compares TCFIs obtained from simulations of 512 and 1000 molecules of the benzene - ethanol mixture at $x_1 = 0.9$. The uncertainties are as expected smaller with 1000 molecules, but the final results do not differ significantly. Consequently when systems of equivalent molecule sizes are studied, 512 molecules will be used.

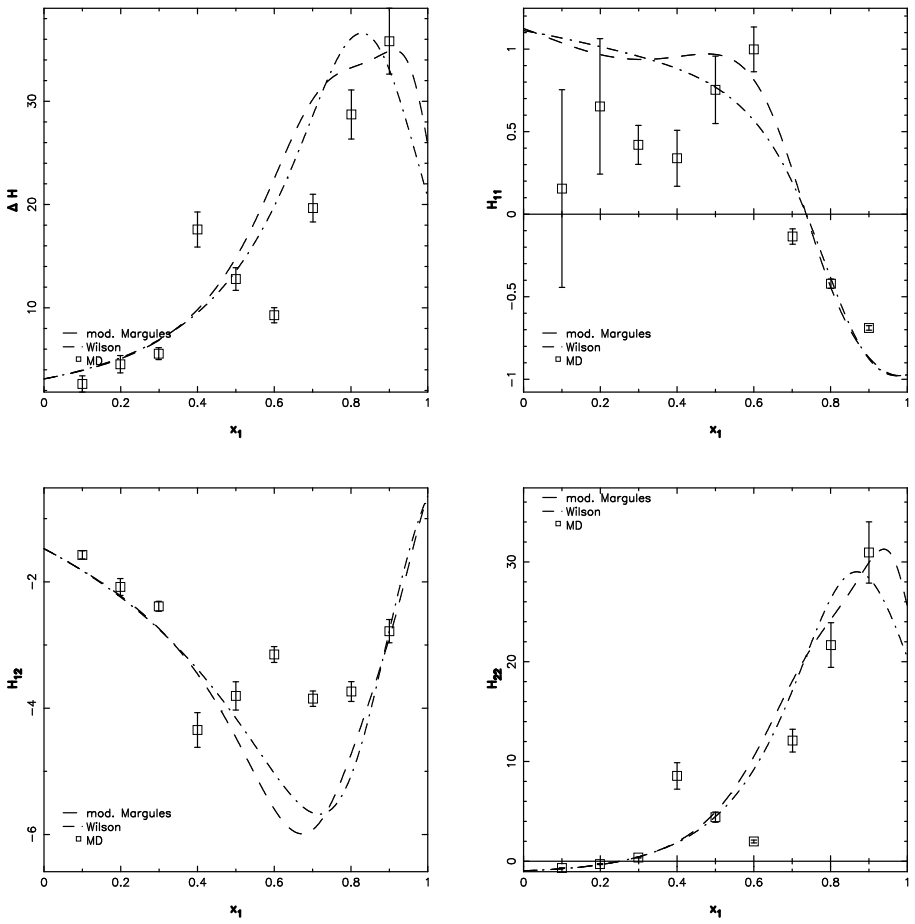


Figure 7.12. Comparison of TCFIs from MD simulations and reverse approach for benzene - ethanol system at 298.15 K. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dot-dashed: Wilson.

Number of molecules	Integration method	H_{11}	H_{12}	H_{22}
512	1	-0.595±0.011	-2.29±0.18	30.9±3.8
	3	-0.689±0.016	-2.78±0.23	···
1000	1	-0.6631±0.0091	-2.317±0.0081	29.8±1.1
	3	-0.8333±0.0059	-2.851±0.043	···

Table 7.8. Comparing TCFIs for benzeneÜethanol system at $x_B = 0.9$ using methods 1 and 3 for integration in two setups with 512 and 1000 molecules, respectively.

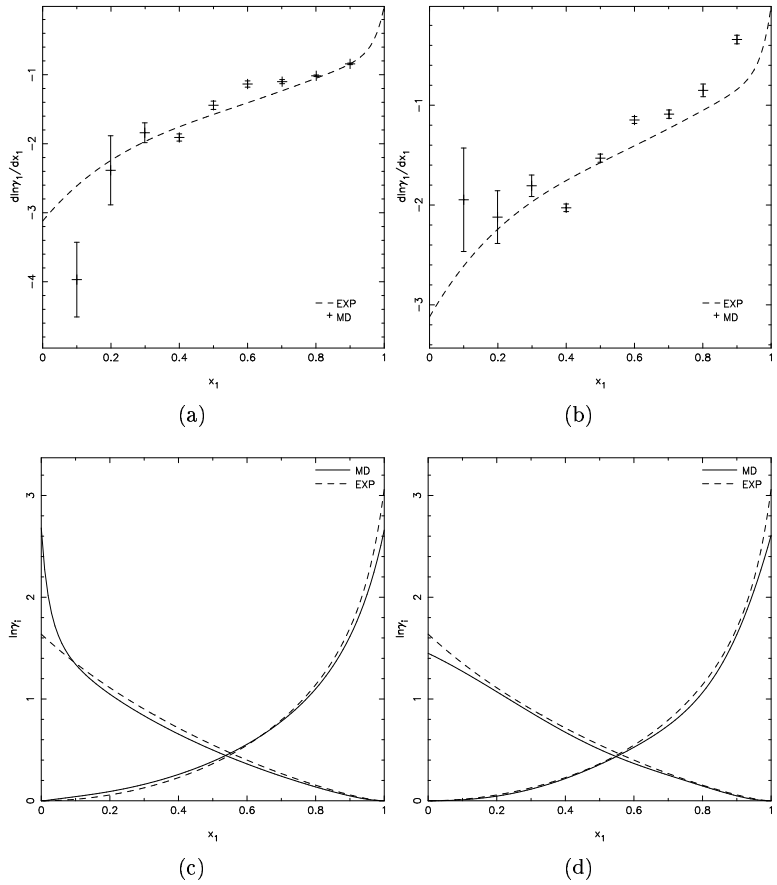


Figure 7.13. Benzene(1) - ethanol(2) at $298.15K$. Comparing results of MD results with modified Margules correlation. In Figures a and c integration method 1 was applied and method 3 was applied in Figures b and d. $*$ Derivatives generated using TCFI. $-$ modified Margules correlation reproduction simulation results. $--$ modified Margules correlation reproducing experimental data Wooley and O'Connell (1991).

The derivatives of the activity coefficient plots in Figure 7.13 show similar results for the method 1 and 3 of the integration methods of the RDFs. The increased uncertainty of applying integration method 1 does not play an important role in this example, possible because the uncertainties are relative small compared to the TCFIs. In Table 7.9 the modified Margules parameter sets op-

	Method 1	Method 3
A_{12}	1.867±0.032	1.442±0.027
A_{21}	2.117±0.091	2.60±0.20
α_{12}	1.82±0.24	0.50±0.28
α_{21}	0.80±0.18	2.2±1.3
η	···	-1.75±0.51

Table 7.9. Modified Margules parameters for benzene (1) - ethanol (2) mixture at 303.15 K determined from FST analysis using method 1 and 3 for integration of RDFs.

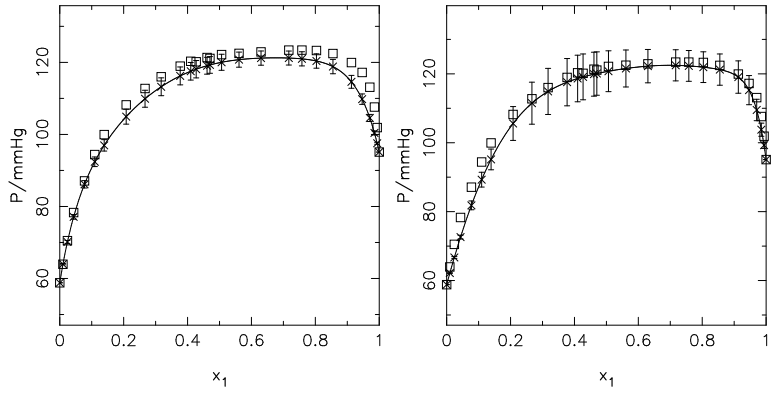
timized from the integrals of method 1 and 3 are listed. The obvious difference is the number of parameters included in the optimal solution for integration method 1 and 3 when following the optimization procedure described in Section 4.1. In both cases there was a significant reduction in the value of the objective function when going from 2 to 4 parameters and also when going from 4 to 5 parameters. For method 1 the uncertainties of the α_{12}, α_{21} and η becomes too significant for the 5 parameter version and it is discarded. For method 3 the uncertainties of α_{12} and α_{21} are of the same order of magnitude as the parameters themselves in the 5 parameter version, however the uncertainties are much larger for α_{12} and α_{21} in the 4 parameter version and consequently the 5 parameter version is selected.

modified Margules parameter set	SCORE	$x_{1,\text{azeotrope}}$
method 1	2.64	0.677
method 3	2.29	0.698
Wooley and O'Connell (1991)	1.81	0.688
(Hwang and Robinson, 1977)	0.3694	0.695

Table 7.10. SCORES (Eq. 5.1) and azeotropic point for benzene - ethanol system at 298.15K where method 1 and 3 had been used for the integration of RDFs in the FST analysis. Compared with two modified Margules parameter set determined from experimental data. The first is from the work of Wooley and O'Connell (1991) have applied another data set in the fitting process and the second is determined using only the experimental data set applied in the evaluation (Hwang and Robinson, 1977).

The modified Margules parameter sets from the FST analysis is evaluated against experimental data of (Hwang and Robinson, 1977) and the results vi-

sualized in Figure 7.14 where the behavior of the system is reproduced by both parameter sets. The system has an azeotropic point and the prediction of the point is listed in Table 7.10 together with scores. Wooley and O'Connell (1991) have treated the system and a modified Margules parameter set was available from their work. When testing the modified Margules description of Wooley and O'Connell (1991) against the experimental data a non-perfect score of 1.81 was obtained. The experimental data applied in the determiation of the parameter set by Wooley and O'Connell (1991) was not retrieved. To test if the data set was inconsistent an additional modified Margules parameter set was optimized using the data of (Hwang and Robinson, 1977). The score of the new parameter set was 0.3694 and the data set is consequently considered consistent. Other approaches for testing consistency require vapor composition data which was not available and the simple approach described above was preferred. In both the case of method 1 and 3 the prediction of the azeotropic point is impressive as the predicted azeotropic points are within a few percent of the experimental determined value. The results show that the FST analysis methodology can detect azeotropes and estimate the azeotropic composition, $x_{\text{azeotrope}}$, with high accuracy for non-ideal systems.



(a) method 1 (b) method 3

Figure 7.14. Bubble point pressure curves for the benzene (1) - ethanol (2) system at 298.15K using 2 sets of modified Margules parameters where method 1 and method 3 has been applied for integration of RDFs in FST analysis.

7.4 Methyl acetate - n-pentane

The methyl acetate (1) - n-pentane (2) system is like the previous system moderately non-ideal and has an azeotropic point. The temperature is 298.15 K and the pressures are still below 1 atm (760 mmHg). The CHARMM descriptions were taken from literature (A. D. MacKerell *et al.*, 1998) and the atomic

descriptions are shown in Figure 7.15. MD simulations were performed for 9

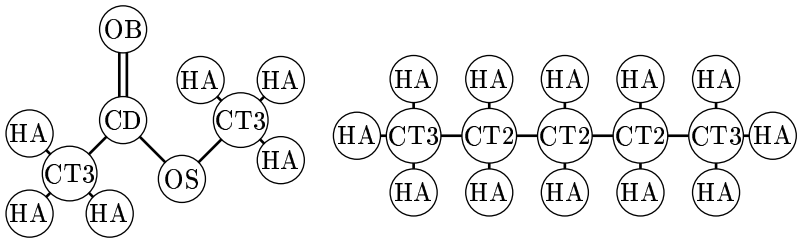


Figure 7.15. Atomistic CHARMM description of methyl acetate and n-pentane

compositions across the composition range. 512 molecules used for the mixtures and standard settings from section 3.2 were applied. Pure component simulations were performed to predict the molar volumes, V_m^{MD}, to evaluate the CHARMM force field description of components. The results are listed in

compound	$V_m^{exp}/\mathrm{cm}^3/\mathrm{mol}$	$V_m^{MD}/\mathrm{cm}^3/\mathrm{mol}$	rel. error %
methyl acetate	79.82	81.0±1.2	1.4±1.5
n-pentane	115.255	120.2±1.9	4.3± 1.7

Table 7.11. Comparing pure compound molar volumes for methyl acetate and n-pentane at 298.15 K.

Table 7.11 and V_m^{MD} of methyl acetate is within the 2% considered acceptable. The molar volume of n-pentane is too high and the relative error could be critical as the RDF and TCFI depend on molar volume. The large inaccuracy is puzzling as data of n-pentane has been included in the optimization of the CHARMM parameters (A. D. MacKerell *et al.*, 1998) where it was the largest of the alkanes.

H_{11} is overestimated at low x_1 by more than 50% in some cases (Figure 7.16). The cause could be the general problem of quality sampling at dilute conditions. Another plausible reason cause is the poor reproduction pure n-pentane molar volume. For H_{12} the agreement between the MD and reverse approach results is very good for the whole composition range and the minimum around $x_1 = 0.5 - 0.6$ is reproduced with acceptable accuracy. H_{22} is underestimated at high methyl acetate concentration, however the trend of the MD simulations match the reverse approach trend. The only exception is at $x_1 = 0.9$ where H_{22} decrease but the uncertainty of H_{22} is relatively large and too much emphasis should not be put into this data point. The resulting ΔH-curve follows the trend of the reverse approach results. It is overestimated at low x_1 and

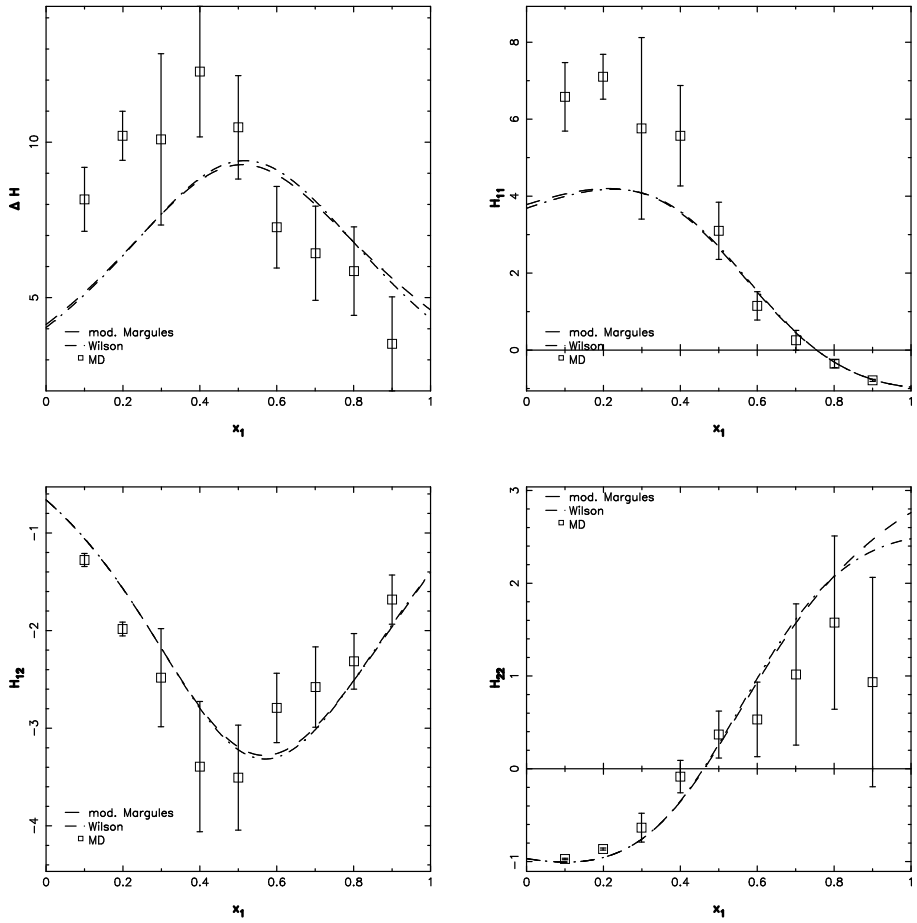
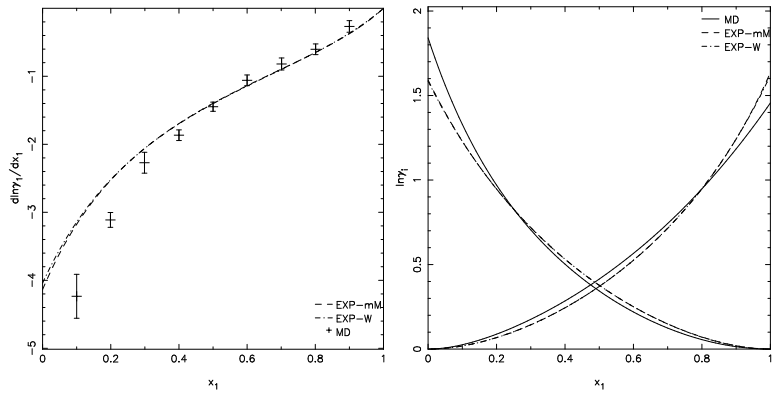


Figure 7.16. Total Correlation Function Integrals (TCFIs) for methyl acetate (1) - n-pentane (2) system at 298.15 K. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dot-dashed: Wilson.

underestimated at high x_1 which is a consequence of the H_{11} and H_{22} errors.



(a) (b)

Figure 7.17. Methyl acetate(1) - n-pentane at $298.15K$ applying integrals from method 3.

The derivatives of the activity coefficient is shown in Figure 7.17.a where the underestimation of ΔH at high methyl acetate concentration is damped when converted to $\partial \ln \gamma_1 / \partial x_1$ because of the mathematical behavior of the Kirkwood-Buff expression (Equation A.35). The overestimation of ΔH at low methyl acetate concentrations results in too negative values of the derivative when compared to the curves generated from experimental data. The resulting $\ln \gamma_1^{MD}$-curve in Figure 7.17.b does not suffer significantly from inaccuracy of $\partial \ln \gamma_1 / \partial x_1$ at low x_1 and the agreement between the γ-curves from FST analysis and experiments is good.

The results from integration method 3 are in general more accurate and further improvement of these might be possible by applying another objective function. Both objective functions, SS_1 and SS_2 (Equations 4.2 and 4.3) have been applied for the optimization of modified Margules descriptions. The resulting parameter sets are very similar and the parameters are listed in Table 7.12. Large reductions were made in both cases when advancing from 2 to 4 parameters in the objective function values, and the expansion with the fifth parameter did not give significant better results and the 4 parameter description was selected in both cases.

Integration method 1 was applied again to investigate if the good results of the benzene - ethanol was coincidental or if general conclusion can be extracted from the results. The modified Margules parameter set obtained is listed in Table 7.12 where the optimal parameter set consisted of 4 parameters which differ from the parameter sets from the other approaches. The result of the difference is negative and the score is significantly larger than the results of the other parameter sets which are equivalent (see Table 7.13). The score from

	Method 1	Method 3	
	SS1	SS1	SS2
A_{12}	2.495±0.057	1.841±0.027	1.803±0.029
A_{21}	1.570±0.050	1.456±0.076	1.515±0.040
α_{12}	4.93±0.89	0.86±0.30	0.45±0.14
α_{21}	0.913±0.046	0.38±0.15	0.46±0.17

Table 7.12. Modified Margules parameters for methyl acetate (1) - n-pentane (2) at 298.15 K, determined from FST analysis using objective functions SS1 (Eq. 4.2) and SS2 (Eq. 4.3) for results of integration method 1 and 3.

integration method 1 is however within reasonable limits and the parameter set can applied in an initial design phases. The scores of the two parameter sets originating from method 3 integration are both very good and the result resemble that of Table 7.10 for the benzene - ethanol system.

Integration method	objective function	SCORE	$x_{\text{azeotrope}}$
1	SS1	5.03	0.207
3	SS1	1.91	0.205
	SS2	2.07	0.213
Exp.		$2.3 \cdot 10^{-5}$	0.205

Table 7.13. SCORES (Eq. 5.1) and azeotropic point for methyl acetate - n-pentane system at 298.15K where method 1 and 3 had been used for the integration of RDFs in the FST analysis and two objective functions have been applied results of integration method 3.

The prediction of the azeotropic points is within 1% which is similar to the results of the benzene - ethanol system. This sparse foundation shows a trend in the results. When the non-ideality of the system reach a certain level the FST analysis becomes very precise.

7.5 Ethanol - (2H)-Heptafluoropropane

The system was studied for competing in the third Industrial Fluid Phase Simulation Challenge (IFPSC). The challenge was to predict the bubble point pressure data (xP) for the 343.13 K-isotherm. All published pure component knowledge was allowed to be applied together with xP-data at 283.17 K. There are three objectives in this case. The first is to study if the simple approach (Section 6.3) for optimization of missing force field parameters is sufficient when the resulting parameter description is applied in MD simulations for FST analysis. The second objective is to study the effect of different assumption

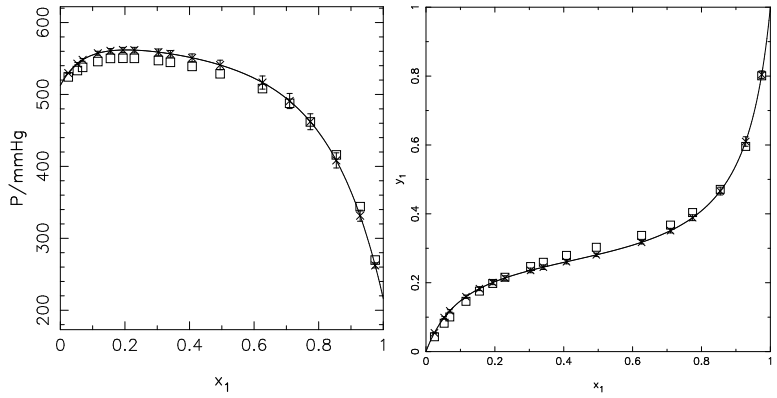


Figure 7.18. Bubble point pressure curve and vapor phase composition as function of liquid phase composition for methyl acetate(1) - n-pentane at $298.15K$. □ is experimental data of Lu *et al.* (1990). Solid lines and × are predictions by the modified Margules parameter set determined using method 3 for integration of RDFs and objective function SS1.

regarding the vapor phase. In previous case studies the selected temperature has been far from the critical temperature, T_c, of any of the components, however in the case at hand T_c of (2H)-Heptafluoropropane is approximately 30 K higher than the studied temperature which together with the vapor pressure of (2H)-Heptafluoropropane indicates that the vapor phase may not be ideal. The third was to test the methodology against world-class investigators in a truly predictive mode.

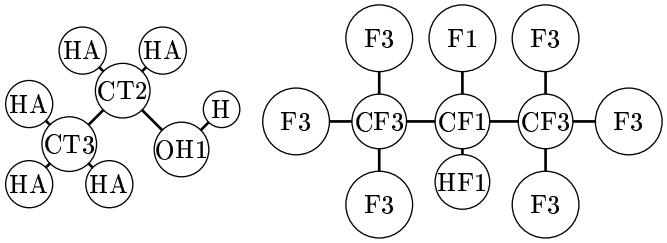


Figure 7.19. Atomistic CHARMM description of ethanol and (2H)-heptafluoropropane.

The CHARMM description of ethanol applied in previous case studies is

reused (See Figure 7.19). No previous CHARMM simulations of (2H)-Heptafluoropropane (HFF-227ea) exist, known to us. Fluorinated ethanes have been simulated by Chen *et al.* (2002) (1-fluoroethane,1,1-difluoroethane, and 1,1,1-trifluoroethane) and by Liu *et al.* (2004) (hexafluoroethane and halothane). Both groups fitted parameters to liquid density and enthalpy of vaporization data, reporting agreements with experimental data within 2%. However, their parameters were optimized only for limited substances and differences appeared for certain atoms. Therefore, the transferability of their parameters to related compounds such as HFF-227ea is questionable. Densities from MD-NPT simulations compared with experimental values (Salvi-Narkhede *et al.*, 1992) showed that the parameter set of Chen *et al.* (2002) gave too low densities, while the set of Liu *et al.* (2004) gave too high densities. The parameter set of Chen *et al.* (2002) was applied with the exception of $R_{min,F3}$. Predictions of HFF-227ea liquid density were most sensitive to that parameter. As a consequence $R_{min,F3}$ was used to fit HFF-227ea liquid densities, using the simple approach described in Section 6.3. The fitting to density data (Table 7.14) gave a value of

T	ρ_{exp}	ρ_{calc}	relative
K	g/mL	g/mL	error
257.55	1.5216	1.55±0.02	-1.92 %
273.15	1.4882	1.50±0.02	-0.76 %
303.07	1.3721	1.39±0.03	-1.56 %
343.38	1.1887	1.21±0.04	-2.10 %

Table 7.14. Density prediction for (2H)-Heptafluoropropane from MD simulation using CHARMM force field.

$R_{min,F3}$ near previously published values. The predicted HFF-227ea densities are typically within 2% of corresponding experimental values and the disagreement never exceeds the uncertainty of the prediction. The atomic charges are the same as those of Chen *et al.* (2002) except for the carbon atoms attached to fluorines. In the united atom force fields available for fluoroalkanes (Cui *et al.*, 1998; Zhang and Ilja Siepmann, 2005), the CF2 and CF3 beads are electro-neutral. This feature was retained by making the corresponding groups of atoms electro-neutral in the HFF-227ea description. Many intra-molecular parameters for HFF-227ea are not available from the literature, so estimates were used, based on similar structures. The description was tested and found acceptable. This was expected, since the intra-molecular potentials are relatively less important to the bulk properties than the intermolecular potentials.

7.5.1 Mixture Simulations

Simulation series were performed at both temperatures and the procedure (Section 3.2) was followed again except at 283.17 K where only 2 production runs were made because of lack of time, but the predictions at 283.15 K were only

made to test the force field description. The modified Margules parameters are listed in Table 7.15 where the parameters generated using integration method 3 for the RDFs and the SS_1 objetive function were applied in the submitted results (Christensen *et al.*, in pressb). The reduction in the number of production runs at 283.17 K is clearly seen in the standard deviations of the parameters where they are a factor 10 higher than what is seen at 343.13 K or in similar systems. One should not compromise with the number of production runs unless the lack of time prohibits the normal set of simulations. The resulting

	Method 1	Method 3	
	SS1	SS1	SS2
$T = 283.17K$			
A_{12}	3.10±0.28	2.31±0.15	1.881±0.054
A_{21}	1.12±0.10	0.75±0.19	-0.31±0.30
α_{12}	10.7±5.6	2.1±1.8	···
α_{21}	1.31±0.20	0.31±0.34	···
η	···	-1.67±0.47	···
$T = 343.13K$			
A_{12}	2.166±0.028	1.818±0.018	1.885±0.074
A_{21}	1.383±0.046	1.227±0.037	1.298±0.014
α_{12}	2.57±0.35	1.57±0.18	2.43±0.86
α_{21}	0.678±0.051	0.510±0.054	0.81±0.16
η	···	···	1.48±0.80

Table 7.15. Modified Margules parameters for ethanol (1) - (2H)-Heptafluoropropane (2) mixture at 283.17 K and 343.13 K determined from FST analysis using objective functions SS1 (Eq. 4.2) and SS2 (Eq. 4.3) for results of integration method 1 and 3.

parameters listed in Table 7.15 for the system at 343.13 K are not the same as those applied in the Challenge problem because the methodology for selecting the best parameter set had not been developed at the time (Christensen *et al.*, in pressb,i). In the entering proposal for the challenge the 5-parameter modified Margules expression was selected.

7.5.2 Vapor Phase Options

Three scenarios have been tested for the vapor phase which are ideal gas, ideal solution and real solution (See description in Chapter 5). Poynting correction of the liquid phase fugacity was neglected.

Assuming ideal gas behavior the results at 283.17 K are in acceptable agreement with experiment giving a SCORE = 6.77 and the Px-diagram is shown in Figure 7.20. At the 343.13 K the SCORE when assuming ideal gas behavior is 1.51 using the 4-parameter description, while the 5-parameter description (submitted in the Challenge) has a SCORE of 1.52 which is equally good.

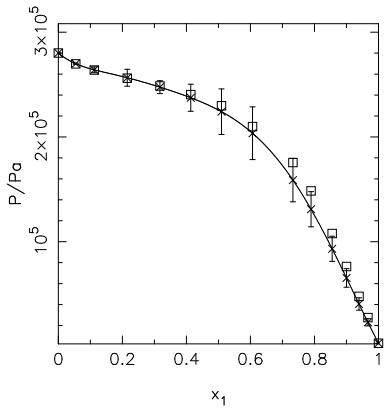
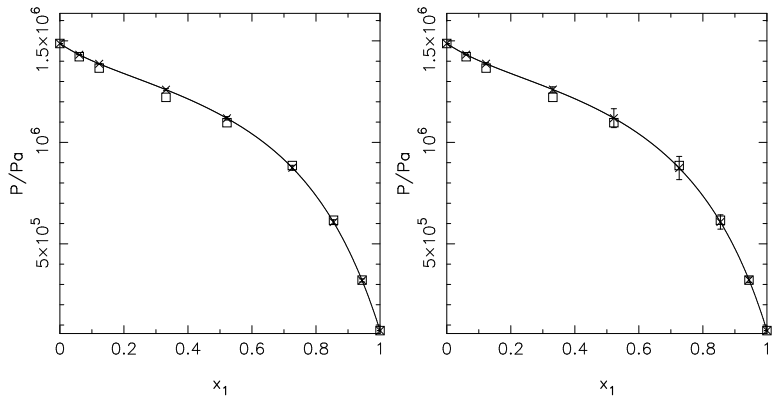


Figure 7.20. Bubble Point pressure curve for ethanol (1) - (2H)-heptafluoropropane (2) at 283.17 K assuming ideal gas behavior.

The difference between the parameter sets is the uncertainty of the parameters which is seen in the pressure predictions in Figure 7.21. The activity coeffi-



(a) 4-parameter description (b) 5-parameter description

Figure 7.21. Bubble Point pressure curve for ethanol (1) - (2H)-heptafluoropropane (2) at 343.13 K assuming ideal gas behavior applying 2 different modified Margules descriptions.

cients found for this system are unusual in behavior, as shown in Figure 7.22. The deviation from Raoult's Law is positive, but there is a maximum for the HFF-227ae and a minimum for the ethanol in the range of x_1 between 0.45 and 0.65.

The method of Hayden and O'Connell (1975) has been used to estimate the second virial coefficients, B_{ij}. The method uses only critical properties

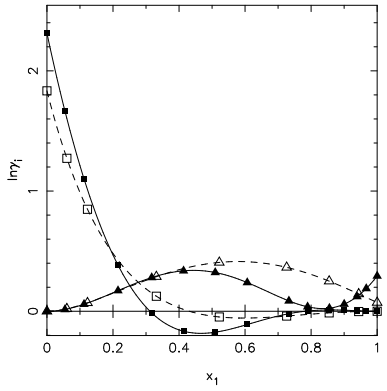


Figure 7.22. Activity coefficients (based on simulations) of ethanol (□) and HFF-227ae (△) at 283.17K (full lines) and 343.13 K (dashes).

(temperature, T_c, and pressure, P_c) and other molecular parameters (radius of gyration, R_G, dipole moment, μ, and solvation parameter, η) as shown in Table 7.16. In the absence of experimental values, all may usually be estimated

Compound	T_c/K	P_c/K	R	μ	η
Ethanol	516.3	63.000	2.25	1.690	1.4
HFF-227ea	374.83	29.116	3.70	1.769	0.0

Table 7.16. Pure component data of ethanol and HFF-227ea applied in method of Hayden and O'Connell (1975) to predict second virial coefficients, B_{ij}.

from molecular structure. We expect no association of C_3HF_7 and, in the absence of direct measurements, no solvation of it with alcohol. A test of this assertion could be made with any HFC compound with an alcohol, since the correlation uses group association/solvation parameters. Parameters based on the radius of gyration are estimated to be only slightly smaller than those of C_3F_8 based on increments for CH_4, C_2H_6, C_3H_8 along with CHF_3, CHF_5, etc. The parameter values used in our calculations are listed in Table 7.16, while the B_{ij} values are given in Table 7.17. The uncertainties in these virial coefficients are probably of the order $\pm$ 100 cm^3/mol. For ethanol, data compilations list values within this range (Dymond and Smith, 1980). For HFF-227ea and the cross coefficients, no information is available for comparison. Figures 7.23 and 7.24 show the effects of vapor non-ideality on P at 283.17 and 343.13K, respectively. There are significant differences, including for y_1 (not shown), indicating that assumptions unrelated to the simulations can affect the results. Tables 7.18 and 7.19 give scores for the various non-ideality approximations and numbers of parameters in the modified Margules model. The ideal gas (IG, where $\Phi_i = 1$) assumption gives somewhat more accurate results for P, even

T/K	B_{total}/cm^3mol^{-1}	B_{free}/cm^3mol^{-1}	B_{chem}/cm^3mol^{-1}
ethanol			
343.15	-1046.03	-79.42	-623.22
283.17	-2963.18	-149.24	-2281.89
HFF-227ea			
343.15	-443.64	-15.84	0.00
283.17	-741.97	-84.43	0.00
mixture, 12			
343.15	-431.53	-48.69	0.00
283.17	-707.09	-116.41	0.00

Table 7.17. From correlation of Hayden and O'Connell (1975), where data from Table 7.16 where applied together with E_{EtOH} = 345.10 K, σ_{EtOH} = 4.710 A, $E_{C_3HF_7}$ = 341.94 K, and $\sigma_{C_3HF_7}$ = 5.158 A. The mixture parameters were E_{12} = 343.52 K, σ_{12} = 4.933 A, and η_{12}= 0.00.

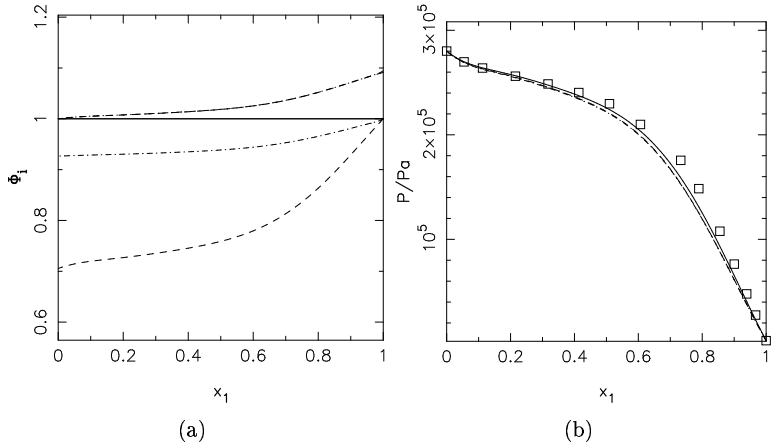


Figure 7.23. Effects of vapor non-ideality and pressure on calculations of P (Pa) at 283.17 K. Ideal Gas ($\Phi_i = 1$): Full curves; Ideal vapor solution ($\delta_{12} = 0$): Dashes; Real vapor solution: Dash-dot curves.

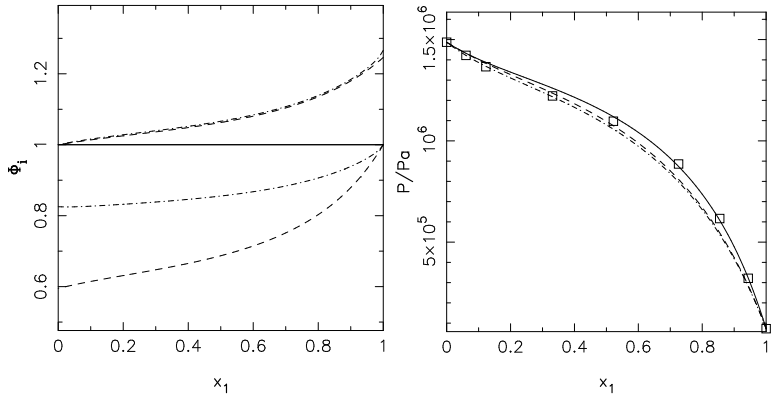


Figure 7.24. Effects of vapor non-ideality and pressure on calculations of P (Pa) at 343.13 K. Ideal Gas ($\Phi_i = 1$): Full curves; Ideal vapor solution ($\delta_{12} = 0$): Dashes; Real vapor solution: Dash-dot curves.

though non-idealities do affect the results. This result is likely to have occurred from the ignoring of non-idealities compensating for errors in the simulations. This effect would probably not be seen for the vapor compositions, providing a more stringent test for the simulations. The scores of Table 7.18 and 7.19 shows that the general accuracy of the methodology is within 10%. VLE calculations (not shown) have also been made for cases in which vapor non-ideality and liquid pressure effects are included. The Poynting correction was included

vapor	mod. Margules		
phase	2	4	5
ig	14.9	5.02	6.77
is	16.8	7.55	9.40
rs	16.8	7.69	9.60

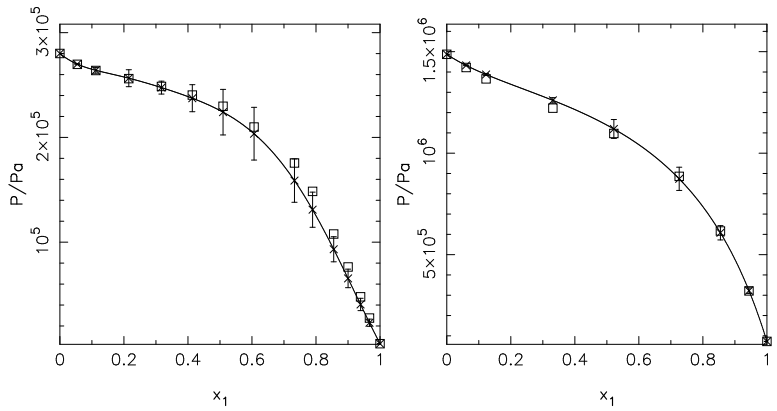
Table 7.18. Score for 2, 4 and 5 parameter modified Margules models and different vapor non-ideality approximatins. IG = ideal gas , IS = ideal solution and RS = real solution at 283.17 K.

as in Equation 5.4 where the liquid volumes were obtained from simulation, as given in Table 7.14. While the effect is small at the lower temperature, it makes several percent difference at the higher T because of the high vapor pressure of HFF-227ae, principally decreasing P at mid-range concentrations.

The predictions submitted for the industrial fluid phase simulation challenge (Figure 7.25) resulted in a score of 1.52 as previously shown. This was the best score, however other research groups and companies did also make good pre-

vapor	mod. Margules		
phase	2	4	5
ig	7.2	1.51	1.52
is	10.0	5.30	5.20
rs	10.2	5.75	5.63

Table 7.19. Score for 2, 4 and 5 parameter modified Margules models and different vapor non-ideality approximatins. IG = ideal gas , IS = ideal solution and RS = real solution at 343.15 K.



(a) T = 283.17 K (b) T = 313.13 K

Figure 7.25. Bubble point pressure for mixture of Ethanol (1) - (2H)-heptafluoro-propane (2) at 283.17 and 343.13 K. Submitted predictions for the third Industriel Pluid Phase Simulation Challenge - state transferability problem. solid line and $\times$ are the predictions and $\square$ are experimental data.

dictions and the top five is given in Table 7.20. The individual methods of the

Rank	Group	score
1	This work	1.52
2	University of Stuttgart, Germany	2.83
3	COSMOlogic, Germany	3.44
4	University of Dortmund, Germany	3.72
5	University of Minnessota, USA	4.24

Table 7.20. Top five from the third industrial fluid phase simulation challenge - state transferability problem.

other groups can not be described here, as the IFPSC work was not published at the deadline of this thesis. It is known that the third place contribution was given by Klamt and Eckert using COSMOtherm previously described in Section 2.2. The fifth place contribution was made by Siepmann *et al.* using a method similar to the GEMC work presented in Section 2.1. The scores of all the top five contributions are within 5% of the experimental data and they can all be considered good enough for application in an initial design process. The result of the Challenge shows that the methodology presented here is competitive with the best predictive methods on a worldwide scale.

7.6 Lactones

The work presented in this last case study is an attempt to test the full potential of the work presented in this thesis. The subject is the lactone compound group (cyclic esters), which has not been simulated previously and a force field description has to be developed before the FST analysis work can be employed. The simple approach (Section 6.3) was used and the objective is to study how the force field parameter transferability concept can be combined with FST analysis when studying a novel compound group. The case study is also dedicated to study the level of precision in the predictions of the force field needed before it can be applied in FST analysis work. The developed force field description for the lactone group is applied in the simulation of γ-butyrolactone and δ-valerolactone with benzene at two temperatures.

7.6.1 Force Field Optimization

The objective is to generate a CHARMM force field description of lactones which is transferable, meaning lactones of different ring size can use the same parameter set and the non-ester part can be described by alkane parameters equivalent to the case for linear esters. In the CHARMM force field no parameters existed for cyclic alkanes and the parameters available for linear alkanes were tested for cyclic hexane and pentane and the results are listed in Table

7.21. CT2 is the descriptor of the linear alkane methylene carbon atom (See n-pentane description in Figure 7.15). Both the molar volume, V_m, and enthalpy of vaporization, ΔH_{vap}, where predicted with unacceptable accuracy using the linear alkane parameters at 298.15K and the cyclic alkane description had to be developed. The strategy for improving the prediction of cyclic alkane prop-

			Relative Errors, %			
	ϵ/kcal/mole	R_{min}/A	V_m^{cp}	ΔH_{vap}^{cp}	V_m^{ch}	ΔH_{vap}^{ch}
CT2	0.055	4.35	-8.42	13.78	-12.45	18.39
CC2	0.074	4.210	1.68	4.27	-0.55	-4.51

Table 7.21. Parameters for linear (CT2) and cyclic (CC2) alkanes and their relative errors for prediction of molar volume, V_m, and enthalpy of vaporization, ΔH_{vap}, for cyclopentane (cp) and cyclohexane (ch) at 298.15 K.

erties were based on the assumption of transferability of the intra molecular potential from linear to cyclic alkanes. This assumption has been applied in both the Buckingham exponential-6 potential (Errington and Panagiotopoulos, 1999a) and the Anisotropic United Atom potential by Bourasseau *et al.* (2002a). The second assumption applied is the unchanged behavior of hydrogen thereby keeping the Lennard-Jones parameters of the hydrogen atoms from the linear alkane description. In the last assumption the atomic charges were kept from the linear alkanes. The only parameters left for the optimization were ϵ_{CC2} and σ_{CC2} of the new cyclic methylene carbon atom, CC2. In the objective function the experimental molar volumes and enthalpy of vaporization for cyclopentane and cyclohexane at 298.15K were included. The results of the optimization is shown in Table 7.21 where the relative errors of V_m are below 2% and the relative errors of ΔH_{vap} are below 5%. The results for V_m are acceptable but the reproduction of ΔH_{vap} could be improved. The maximum relative error generally applied in optimization of force fields is 2%. Compared to the work by Errington and Panagiotopoulos (1999a) the present results are acceptable, as their Buckingham exponential-6 parameter description for cyclohexane could not be transfered to cyclopentane.

Further development of the cyclic alkane force field description could be performed to reach relative errors of ΔH_{vap} below 2% where state transferability also could be included. The procedure could include determination of partial charges by atomistic techniques like the restrained electrostatic potential (RESP) method (Bayly *et al.*, 1993; Levy and Enescu, 1998). A possible change of the partial charges could both give better (or worse) predictions with the CC2 parameters of Table 7.21. The optimization of Lennard-Jones parameters could be repeated where only CC2 parameters are included if this does not result in an improvement compared to previous results it is possible to also include the Lennard-Jones parameters of the hydrogen atoms.

The generation of a cyclic alkane has paved the way for the optimization of the cyclic ester or lactone parameters. The transferable parameter description

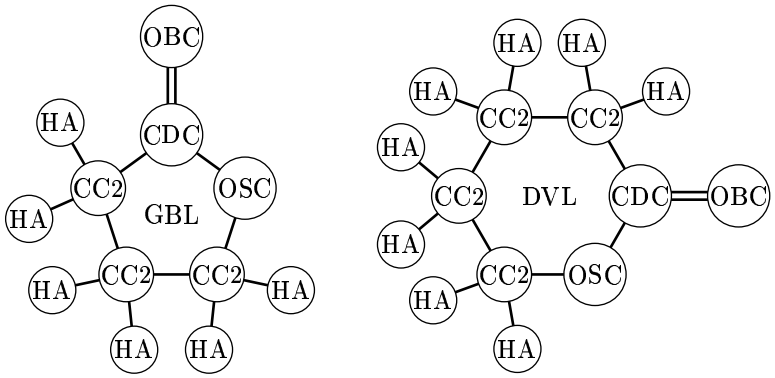


Figure 7.26. Atomistic CHARMM description of γ-butyrolactone (GBL) and δ-valerolactone (DVL).

consists of the ester carbon atom, CDC, the double bonded oxygen, OBC, and the single bonded oxygen, OSC. The intramolecular parameter description is taken from the linear ester description where the descriptors are CD, OB and OS (See methyl acetate in Figure 7.1) and the initial parameters for the lactones are taken from the existing linear ester parameters. Partial charges were taken from the linear ester description. This might create problems but the simple approach is followed (Section 6.3), nevertheless. Missing intramolecular parameters are taken from similar structures from linear alkanes and esters. The most sensitive Lennard-Jones parameters are those of the oxygen atoms, OBC and OSC, which will be included in the optimization.

Experimental data (Table 7.22) show a clear difference in the pure component properties molar volume, V_m, and enthalpy of vaporization, ΔH_{vap}, between equivalent linear esters and lactones. V_m of the linear ester are generally 25% larger than the lactone values. ΔH_{vap} is approximately 15-18 kJ/mol higher for the lactones which relatively is up to 50% higher. This major difference in experimental data has to be reflected in the parameter description. Experimental data of 4 lactones (γ-butyrolactone, δ-valerolactone, γ-valerolactone, and ϵ-caprolactone) were included in the parameter optimization, and molar volume data of γ-butyrolactone were included covering a small temperature range to improve the state transferable of the force field. Two force fields were developed to study how extensive an objective function is necessary for optimization of a force field which is to be applied in binary liquid simulations for FST analysis. In optimization of the first parameter set, only molar volume data were included and a comparison of experimental and predicted data is listed in Table 7.23. The predicted molar volumes are within 2% of the experimental values with the exception of δ-valerolactone. The predicted enthalpies of vaporization are all poorly predicted. This is not inexpected as the experimental values

compound	M_w g/mol	V_m mL/mol	ΔH_{vap} kJ/mol
methyl acetate	74.08	79.82	32.6
ethyl acetate	88.11	98.59	35.1
propyl acetate	102.13	115.76	39.8
isopropyl acetate	102.13	117.58	37.2
butyl acetate	116.16	132.59	43.9
β-propiolactone	72.06	63.07	47.0
γ-butyrolactone	86.09	76.54	53.5
δ-valerolactone	100.12	92.77	58.0
γ-valerolactone	100.12	94.70	54.8
ϵ-caprolactone	114.14	110.80	62.0

Table 7.22. Comparison of molar volume, V_m, and enthalpy of vaporization, ΔH_{vap} for equivalent linear esters and lactones.

Lactone	T/K	$V_m/mL/mol$			$\Delta H_{vap}/kJ/mol$		
		exp	MD	RE/%	exp	MD	RE/%
γ-BL	298.15	76.54	76.60±0.53	-0.1	53.5	37.23	30
	318.16	77.91	77.74±0.57	0.2	···	···	
	338.15	79.32	79.02±0.64	0.4	···	···	
	368.14	81.54	80.91±0.71	0.8	···	···	
γ-VL	298.15	94.70	95.04±0.67	-0.4	54.8	···	
δ-VL	298.15	92.77	90.64±0.6	2.3	58.0	35.4	39
ϵ-CL	298.15	110.80	109.65±0.73	1.0	62.0	33.7	46

Table 7.23. Comparison of molar volume and enthalpy of vaporization for lactones for parameter set 1 with experimental data. γ-BL: γ-butyrolactone, γ-VL: γ-valerolactone, δ-VL: δ-valerolactone, and ϵ-CL: ϵ-caprolactone.

were not included in the objective fuunction. In the optimization of the sec-

Lactone	T/K	$V_m/mL/mol$			$\Delta H_{vap}/kJ/mol$		
		exp	MD	RE/%	exp	MD	RE/%
γ-BL	298.15	76.54	78.09±0.38	-2.0	53.5	63.7	-19
	318.16	77.91	78.82±0.41	-1.2	···	···	
	338.15	79.32	79.73±0.44	-0.5	···	···	
	368.14	81.54	80.95±0.48	0.7	···	···	
γ-VL	298.15	94.70	95.08±0.50	-0.4	54.8	40.3	27
δ-VL	298.15	92.77	91.89±0.44	1.0	58.0	57.8	0.4
ϵ-LC	298.15	110.80	109.54±0.51	1.1	62.0	55	12

Table 7.24. Comparison of molar volume and enthalpy of vaporization for lactones for parameter set 2 with experimental data. γ-BL: γ-butyrolactone, γ-VL: γ-valerolactone, δ-VL: δ-valerolactone, and ϵ-CL: ϵ-caprolactone.

ond parameter set ΔH_{vap}-data were include in the objective function and the results are listed in Table 7.24. The predicted molar volumes are within 2% of the experimental values, however it was not possible to determine a parameter set where ΔH_{vap} was predicted within 2%. Thus, the results are not entirely satisfactory, but they are better than those found with parameter set 1.

There are several possible sources of problem with reproducing the enthalpies of vaporization. One source could be error propagation from the cyclic alkane parameters where the accuracy of the ΔH_{vap} predictions are only 5%. A second source could be the reuse of the partial charges from the linear esters and alkanes. The partial charges could be a good starting point for improving the intermolecular part of the force field.

The optimized parameters are listed in Table 7.25 where the most significant change is the ϵ-parameters. The difference in molar volume between linear ester and lactones were reproduced by a combination of modifying both OSC and OBC parameters for parameter set 1. In the optimization it was attempted to keep the parameters of the oxygen atoms within range of each other, similar to the linear ester parameter relations. In order to reproduce ΔH_{vap} of the

Parameter		init.	par. set 1	par. set 2
OBC (double bonded)				
ϵ	kcal/mol	-0.120	-0.350	-0.640
R_{min}	A	3.4	3.29	3.76
OSC (single bonded)				
ϵ	kcal/mol	-0.152	-0.300	-0.650
R_{min}	A	3.54	3.30	3.80

Table 7.25. Optimized Lennard-Jones parameters 1 and 2 for lactone force field.

lactones in the parameter set 2, it was necessary to increase ϵ_{OBC} and ϵ_{OSC} by a factor 4-5 and they are significantly larger than any published ϵ-parameters. This suggests that other steps might be worthwhile considering. Here the consequences are studied by applying both parameter set 1 and 2 to mixtures of both γ-butyrolactone and δ-valerolactone with benzene. The description of benzene is reused from previous cases (see Figure 7.1 or 7.10) the lactone descriptions are shown in Figure 7.26.

7.6.2 Lactones with benzene

The two parameter sets have been applied to γ-butyrolactone - benzene at 293.15K and δ-valerolactone - benzene at 293.15K and 313.15K. This has been done to test both the transferability of the parameter description to different lactones and secondly to test the state transferability capabilities.

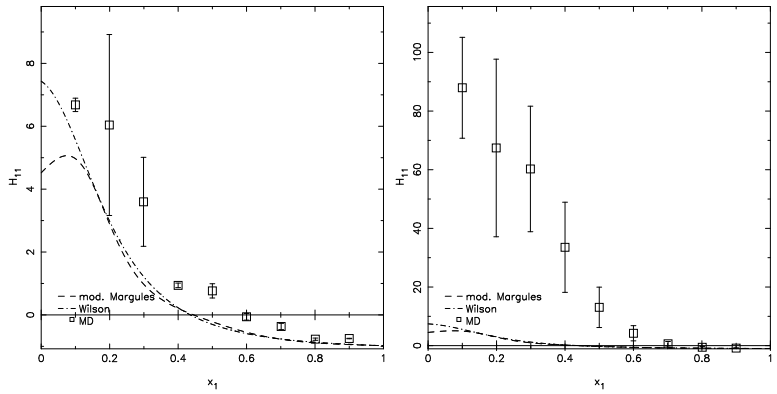
For all cases MD simulations were performed for 9 compositions across the composition range. 512 molecules used for the mixtures and standard settings from Section 3.2 were applied.

7.6.2.1 γ-butyrolactone systems

The TCFIs of simulations using both parameter sets are compared with each other and reverse approach results. For the Huang-O'Connell correlation for κ_T no parameter description was available for γ-butyrolactone in the original work of Huang and O'Connell (1987). The three parameters C^*, V^* and T^* where optimize to reproduce experimental pure component density-temperature-pressure data of Ihmels and Gmehling (2002) and the parameters are listed in Table D.1.

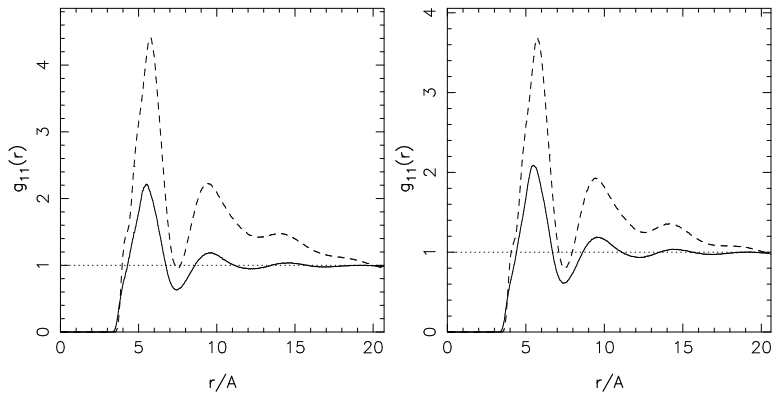
In Figure 7.27.a the reproduction of H_{11} from reverse approach calculations is good for parameter set 1, though with some disagreement at lower concentrations of γ-butyrolactone where TCFIs from simulations are overestimated. The results for parameter set 2 are very different, where H_{11} is very big for $x_1 \leq 0.5$ (Fig. 7.27.b). This is caused by self-association, which gives a high first peak in the $g_{11}(r)$. The $g_{11}(r)$ are compared at two compositions in Figure 7.28.

The reproduction of H_{12} in Figure 7.29.a is good in both ends of the composition scale, but the TCFIs from $x_1 = 0.3$ to $x_1 = 0.7$ H_{12} is too negative which indicate less random distribution than the experimental data indicate. This phenomenon is almost taken to the extreme in Figure 7.29.b where the very negative values of H_{12} in the range $x_1 = 0.2 - 0.4$ could indicate a liquid-liquid split. H_{22} is reproduced with similar accuracy of H_{11} and H_{12} for parameter set 1 in Figure 7.30.a with too positive TCFIs at equimolar conditions. The profile of H_{22} for parameter set 2 gives a maximum of 10 at equimolar composition. The benzene-benzene TCFIs from simulations are clearly influenced by the change in the lactone force field, the only agreement is found at the lowest concentration of γ-butyrolactone ($x_1 = 0.1$). This difference between



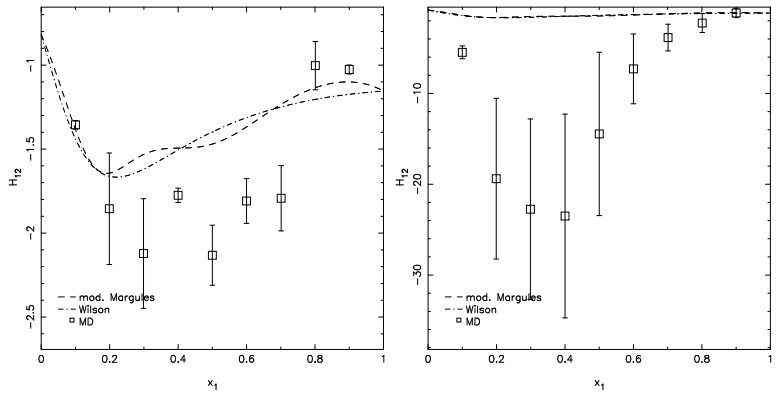
(a) parameter set 1 (b) parameter set 2

Figure 7.27. Comparison of H_{11} from γ-butyrolactone (1) - benzene (2) mixture simulations using parameter set 1 and 2. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dot-dashed: Wilson.



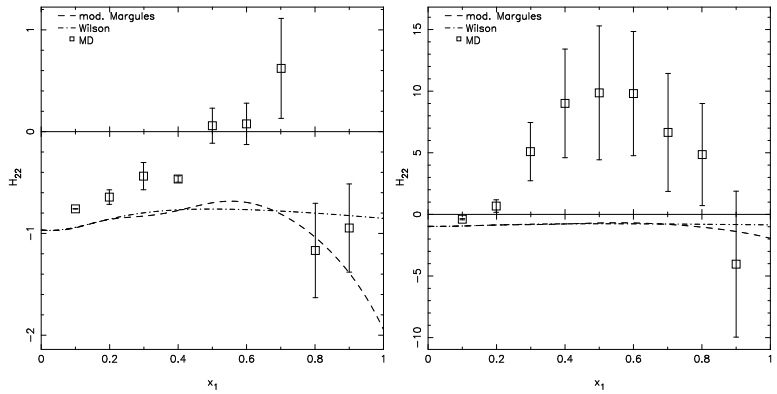
(a) $x_1 = 0.3$ (b) $x_1 = 0.4$

Figure 7.28. Comparison of $g_{11}(r)$ from γ-butyrolactone (1) - benzene (2) mixture simulations at two compositions. solid line: CHARMM parameter set 1, dashed line: CHARMM parameter set 2.



(a) parameter set 1 (b) parameter set 2

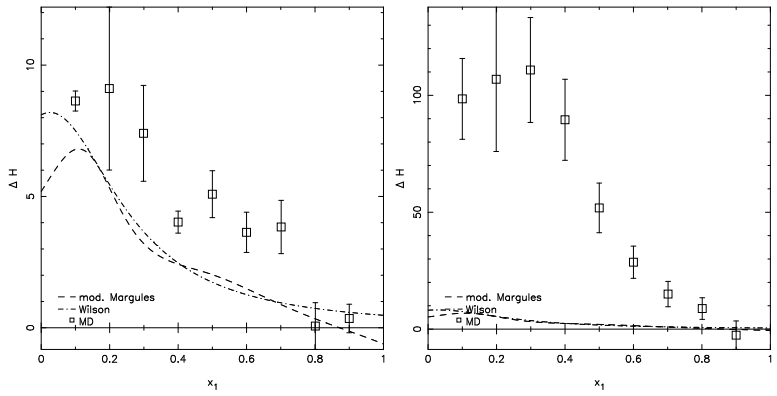
Figure 7.29. Comparison of H_{12} from γ-butyrolactone (1) - benzene (2) mixture simulations using parameter set 1 and 2. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dot-dashed: Wilson.



(a) parameter set 1 (b) parameter set 2

Figure 7.30. Comparison of H_{22} from γ-butyrolactone (1) - benzene (2) mixture simulations using parameter set 1 and 2. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dot-dashed: Wilson.

H_{22} behavior is a direct consequence of the results for H_{11} because the γ-butyrolactone molecules cluster then the benzene molecules are automatically forced to be among themselves. The values of ΔH in Figure 7.31 are gener-



(a) parameter set 1 (b) parameter set 2

Figure 7.31. Comparison of ΔH from γ-butyrolactone (1) - benzene (2) mixture simulations using parameter set 1 and 2. □ are results from simulation. Lines are from reverse approach of Wooley and O'Connell (1991), dashed: modified Margules and dot-dashed: Wilson.

ally too high when compared to the reverse approach. The overestimation at $x_1 = 0.2$ and $x_1 = 0.3$ are caused by the over estimation of H_{11}, but in the range $x_1 = 0.4 - 0.7$ the overestimation is an accumulative difference of the individual TCFIs. For parameter set 2 simulations ΔH is extremely overestimated which for the most part is due to the high values of H_{11} (Fig. 7.27.b).

The different combinations of integration methods of RDFs and objective functions were applied in the optimization of modified Margules parameters for the MD results of both CHARMM parameter sets and the optimal modified Margules parameter descriptions are listed in Table 7.26. For all the results of from CHARMM parameter set 2 the simplest description was chosen as the increase in the number of parameters did not reduce the value of the objective function significantly.

The scores of the modified Margules parameter descriptions are listed in Table 7.27 where the rough reproduction of the TCFIs of CHARMM parameter set 1 is carried forward to the pressure predictions where the scores are in the range 10-20%. An example is shown in Figure 7.32 where deviation from ideal behavior of the liquid phase is overestimation by the modiied Margules description from FST analysis.

As already indicated by the reverse analysis a liquid-liquid split is detected for two of the three modified Margules description determined from simulations

	Method 1	Method 3	
	SS1	SS1	SS2
CHARMM parameter set 1			
A_{12}	1.658±0.077	1.673±0.051	1.568±0.067
A_{21}	0.23±0.26	0.92±0.14	0.72±0.13
α_{12}	···	1.45±0.59	···
α_{21}	···	0.41±0.11	···
CHARMM parameter set 2			
A_{12}	2.46±0.49	2.62±0.28	2.204±0.036
A_{21}	-0.2±1.6	0.15±0.94	0.84±0.29

Table 7.26. Modified Margules parameters for γ-butyrolactone (1) - benzene (2) mixture at 293.15 K determined from FST analysis using objective functions SS1 (Eq. 4.2) and SS2 (Eq. 4.3) for results of integration method 1 and 3. Where two different CHARMM parameter descriptions were applied in MD simulations.

Integration method	Objective function	NO. parameters	SCORE
CHARMM parameter set 1			
1	SS_1	2	12.45
3	SS_1	4	16.27
3	SS_2	2	17.79
CHARMM parameter set 2			
1	SS_1	2	...[a]
3	SS_1	2	...[a]
3	SS_2	2	38.43

Table 7.27. SCORES (Eq. 5.1) for γ-butyrolactone - benzene system at 293.15K using modified Margules descriptions of Table 7.26 and assuming ideal gas phase. [a] liquid-liquid split.

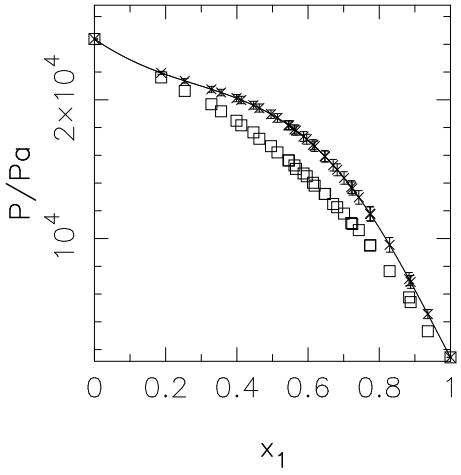


Figure 7.32. Pressure predictions for γ-butyrolactone (1) - benzene (2) system at 293.15K. The modified Margules applied was determined from FST analysis of MD simulation using CHARMM parameter set 1 and the objective function SS_2 (Eq. 4.3). □ are experimental data from Klein and Svejda (1995).

using parameter set 2. In Figure 7.33 the Gibbs energy of mixing, ΔG, has been calculated to exemplify the liquid-liquid split. Similar results are returned by both modified Margules descriptions giving LLE. Even though the last parameter set does not predict LLE, stability calculations show that the mixture is on the verge of splitting into two phases.

7.6.2.2 δ-valerolactone systems

The reverse analysis is not possible as no parameter description was available for the Huang-O'Connell correlation (κ_T) for δ-valerolactone in the original work of Huang and O'Connell (1987). In addition no experimental density-temperature-pressure data could be found in literature. The FST analysis was carried out for the simulation results at both temperatures using CHARMM parameter set 1 and 2. In Figure 7.34 $dln\gamma_1/dx_1$ generated from MD simulations are displayed using integration method 1 and 3 of the RDFs and comparing the results of the two CHARMM parameter sets. The results are quite similar when comparing the integration method 1 and 3. The main reason is the almost complete convergence of the radial distribution functions from the simulations and consequently the only difference is the long distance contribution, H_{ij}^{ld}, which is added in integration method 3. However the long distance contributions are relatively small compared to H_{ij} and become almost insignificant. There is good agreement between the derivative generated from experimental data and the results from applying CHARMM parameter set 1. In the composition range $x_1 = 0.3 - 0.6$ the derivatives from simulation are too low and at

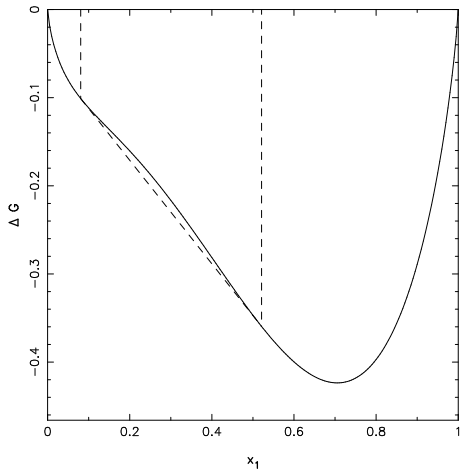


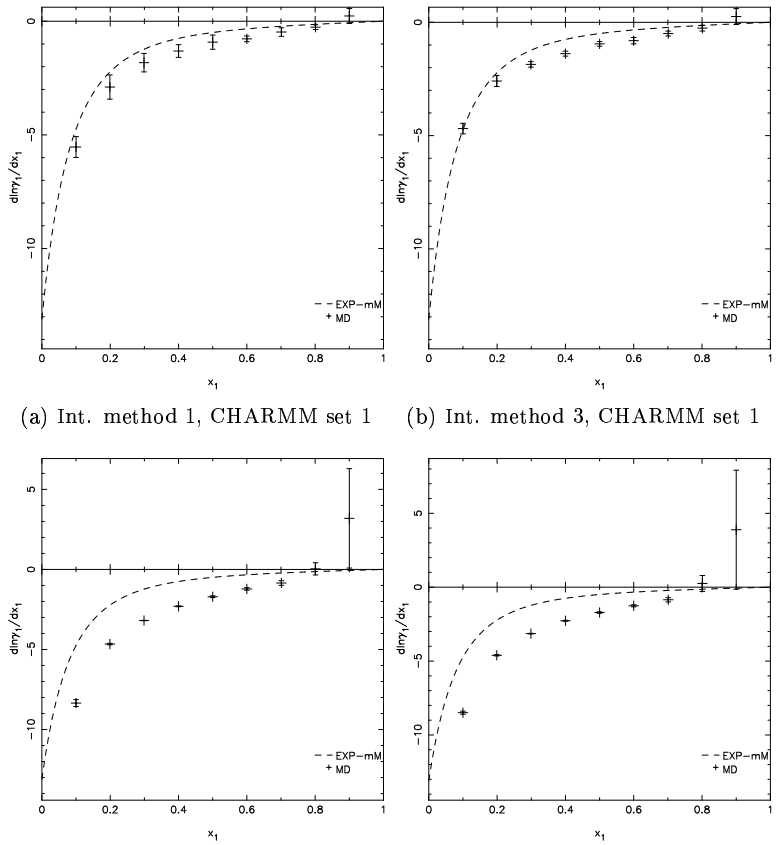
Figure 7.33. Gibbs energy of mixing for γ-butyrolactone (1) - benzene (2) system at 293.15K. The modified Margules applied was determined from FST analysis of MD simulation using CHARMM parameter set 2 and the objective function SS_1 (Eq. 4.2).

$x_1 = 0.9$ a small positive derivative is given where the experimental correlation returns small negative values. For CHARMM parameter set 2, $dln\gamma_1/dx_1$ from simulations have an average error above 100% when compared to the results of the correlation. The modified Margules descriptions generated from the FST analysis at 293.15K are lised in Table 7.28. At 293.15K the results for δ-

	Method 1	Method 3	
	SS1	SS1	SS2
CHARMM parameter set 1			
A_{12}	1.586±0.091	1.45±0.11	1.790±0.093
A_{21}	0.11±0.31	0.24±0.38	0.911±0.060
α_{12}	···	···	3.2±1.5
α_{21}	···	···	0.557±0.099
CHARMM parameter set 2			
A_{12}	2.48±0.27	2.33±0.42	2.11±0.25
A_{21}	0.31±0.92	-0.3±1.4	0.72±0.29

Table 7.28. Modified Margules parameters for δ-valerolactone (1) - benzene (2) mixture at 293.15 K determined from FST analysis using objective functions SS1 (Eq. 4.2) and SS2 (Eq. 4.3) for results of integration method 1 and 3. Where two different CHARMM parameter descriptions were applied in MD simulations.

valerolactone - benzene are similar to those of γ-butyrolactone - benzene where scores in the area of 10-20 are obtained from the FST analysis of the simulations



(a) Int. method 1, CHARMM set 1 (b) Int. method 3, CHARMM set 1

(c) Int. method 1, CHARMM set 2 (d) Int. method 3, CHARMM set 2

Figure 7.34. δ-valerolactone (1) - benzene (2) at 293.15 K. Comparing results of MD results with the modified Margules correlation. Figures a and c uses the TCFIs found using integration method 1. Figures b and d use the TCFIs found using integration method 3. In Figures a and b CHARMM parameter set 1 has been used in simulations and in Figures c and d CHARMM parameter set 2. Crosses and solid lines are results from MD simulations. Dashed lines are generated by modified Margules where the parameters have determined using experimental data (Klein and Svejda, 1995).

Integration method	Objective function	NO. parameters	SCORE
CHARMM parameter set 1			
1	SS_1	2	11.49
3	SS_1	2	9.05
3	SS_2	4	15.23
CHARMM parameter set 2			
1	SS_1	2	...[a]
3	SS_1	2	...[a]
3	SS_2	2	32.52

Table 7.29. Scores (Eq. 5.1) for δ-valerolactone - benzene system at 293.15 K using modified Margules descriptions of Table 7.28 and assuming ideal gas phase. [a] liquid-liquid split.

using CHARMM parameter set 1. In similar fashion does the FST analysis of simulations with CHARMM parameter set 2 return results which can not be applied in any form of predictions for design of equipment for chemical plants.

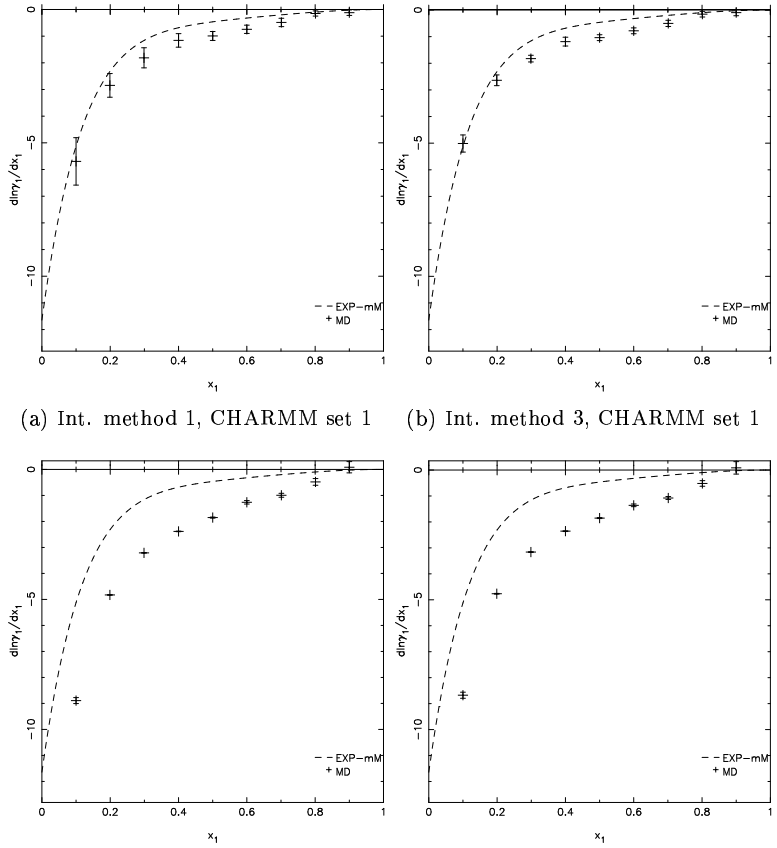
The results at 313.15 K are similar to those at 293.15 K. If one compares Figures 7.34 and 7.35 $dln\gamma_1/dx_1$ from simulations using CHARMM parameter set 1 is too low at midrange compositions, but in general the trend from experimental data is reproduced. In Figure 7.35.c-d the difference between the experimental curves and points from FST analysis of $dln\gamma_1/dx_1$ are similar to those of Figure 7.34.c-d which again shows the CHARMM parameter set 2 does not give reliable results.

The modified Margules parameters are presented in Table 7.30 and the scores of these are presented in Table 7.31 where ideal gas behavior has been assumed. The scores of the mM descriptions from simulations using CHARMM parameter set 1 are equivalent to those at 293.15 K and the scores of γ-butyrolactone - benzene with values above 10%. Opposite to the results at 293.15 K no liquid-liquid split is detected which indicates 313.15 K is above the upper consolute temperature for the results using CHARMM parameter set 2. The scores are enourmous and the modified Margules descriptions generated using CHARMM parameter set 2 are not fit for any use in initial design processes.

An example of the pressure prediction at both temperatures is shown Figure 7.36 where slightly too high pressures are predicted at midrange compositions.

7.6.3 Comment on Lactone Results

None of the binary mixture results presented for lactones show similar accuracy of previously studied systems where CHARMM parameters were taken from the work of A. D. MacKerell *et al.* (1998). This indicates the simple approach is not sufficient when novel compound groups are studied. For novel systems



(a) Int. method 1, CHARMM set 1 (b) Int. method 3, CHARMM set 1

(c) Int. method 1, CHARMM set 2 (d) Int. method 3, CHARMM set 2

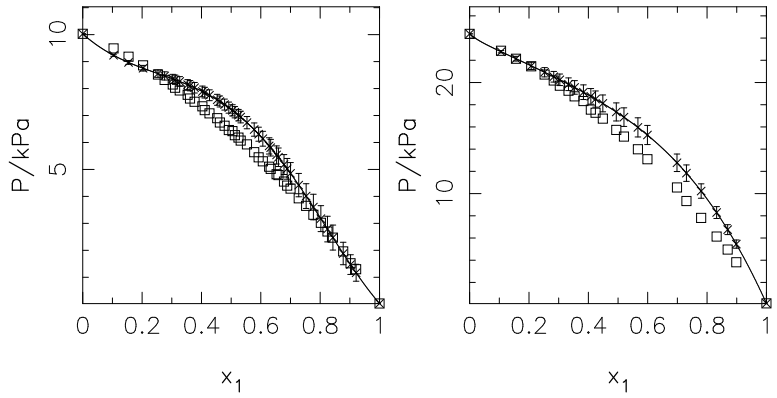
Figure 7.35. δ-valerolactone (1) - benzene (2) at $313.15K$. Comparing results of MD results with modified Margules correlation. Figures a and c uses the TCFIs found using integration method 1. Figures b and d use the TCFIs found using integration method 3. In Figures a and b CHARMM parameter set 1 has been used in simulations and in Figures c and d CHARMM parameter set 2. Crosses and solid lines are results from MD simulations. Dashed lines are generated by modified Margules where the parameters have determined using experimental data (Klein and Svejda, 1995).

	Method 1	Method 3	
	SS1	SS1	SS2
CHARMM parameter set 1			
A_{12}	2.08±0.14	1.97±0.15	1.313±0.048
A_{21}	0.89±0.13	0.885±0.093	0.771±0.091
α_{12}	5.9±2.5	5.9±2.9	···
α_{21}	0.90±0.12	0.76±0.16	···
CHARMM parameter set 2			
A_{12}	3.17±0.22	3.22±0.26	3.084±0.089
A_{21}	1.46±0.14	1.50±0.15	1.826±0.069
α_{12}	6.6±3.8	7.1±4.7	3.58±0.76
α_{21}	1.02±0.14	1.06±0.16	0.77±0.21
η	···	···	-1.34±0.17

Table 7.30. Modified Margules parameters for δ-valerolactone (1) - benzene (2) mixture at 313.15 K determined from FST analysis using objective functions SS1 (Eq. 4.2) and SS2 (Eq. 4.3) for results of integration method 1 and 3. Where two different CHARMM parameter descriptions were applied in MD simulations.

Integration method	Objective function	NO. parameters	SCORE
CHARMM parameter set 1			
1	SS_1	4	13.61
3	SS_1	4	12.97
3	SS_2	2	11.26
CHARMM parameter set 2			
1	SS_1	4	49.40
3	SS_1	4	50.85
3	SS_2	5	60.54

Table 7.31. Scores (Eq. 5.1) for δ-valerolactone - benzene system at 313.15 K using modified Margules descriptions of Table 7.30 and assuming ideal gas phase.



(a) T=293.15 K (b) T=313.15 K

Figure 7.36. Bubble point pressure predictions for δ-valerolactone (1) - benzene (2) system at 293.15 K and 313.15 K. The modified Margules applied was determined from FST analysis of MD simulation using CHARMM parameter set 1 and the objective function SS_1 (Eq. 4.2). □ are experimental data from Klein and Svejda (1995).

a more extensive effort is required both with respect to intramolecular parameterization and determination of atomic charges together with optimization of Lennard-Jones parameters before the parameter description can be applied in the methodology presented here. However it is still an open question how exhaustive the objective function needs to be. Which set of pure component properties should be reproduced accurately by the force field description before it is sufficient to be applied in the methodology presented in this thesis?

8

Conclusion

A methodology has been presented which combines FST analysis with MD simulations to generate a G^E-description for a binary mixture. The only input needed are the pure component force field descriptions of the compounds to be studied. The modified Margules G^E-model description contains standard deviations of the generated parameters which enables uncertainty calculations of bubble point pressures in VLE predictions.

The key quantity generated in the MD simulations is the angle-averaged center-of-mass to center-of-mass radial distribution functions. The RDFs have slow convergence at long distances which can cause divergence of their integrals (TCFIs). Three methods for overcoming that problem have been studied and the best results are obtained when combining numerical integration of short-range RDF with a convergent analytical correlation of the long-range behavior where previous attempts of Matteoli and Mansoori (1995) were improved. The method of Weerasinghe and Smith (2003) where the system size is increased was not found feasible. A strategy for selecting the optimal number of parameters for the modified Margules model has been presented which is similar to earlier work where the data to be reproduced was experimental measurements instead of output from FST analysis.

The FST analysis methodology presented has been applied to several cases to study strengths and weaknesses of the method. The case studies covered a large range of conditions and compound types which shows the versatile nature of the methodology. For systems showing close to ideal behavior, benzene - methyl acetate and methyl acetate - acetone, the pressure predictions were of great accuracy with scores of 1.41 and 2.54, even though thre is a small-difference-of-large-numbers problem near ideality. In the case of methyl acetate - acetone the azeotropic behavior was detected although a double azeotrope was predicted instead of the single azeotrope of the experimental data. The moderately non-ideal systems, benzene - ethanol and methyl acetate - n-pentane, were giving equally good results with scores of 2.29 and 1.91. Both systems contained azeotropic points which were predicted with exceptional precision. For benzene - ethanol the experimental value was 0.695 and the predicted 0.698. In the case of methyl acetate - n-pentane the predicted value was equal to the experimental value. The reverse approach analysis of the benzene - ethanol system show reproduction of ethanol - ethanol self association at dilute conditions by showing high values of the like-like TCFI. The methodology is very accurate

for non-ideal systems, and acceptably accurate for almost ideal systems.

A simple approach for generation of force field descriptions for novel compounds was proposed. The approach was initially applied to (2H)-heptafluoropropane where Lennard-Jones parameters of the dominant F3 molecule were tuned to reproduce liquid densities. The resulting CHARMM description was applied in the system ethanol - (2H)-heptafluoropropane which included reproduction of experimental VLE data at two temperatures, 283.17K and 343.13 K. VaporŰliquid equilibrium predictions have been obtained when vapor nonidealities are ignored (ideal gas), and when estimated second virial coefficients are used in an ideal vapor solution approximation and with a complete formulation. At 283.17K with an ideal gas vapor phase the bubble point pressure (P-x) diagram was predicted with a score of 6.77. At 343.13 K, an ideal gas vapor phase also gave very good results (1.51). Taking vapor non-ideality into account led to lower, but acceptable scores. The results also showed that the presented methodology is very competitive when competing against other methods in prediction of phase behavior.

The simple approach of generation of force field description was finally tested on a wider scale where the lactone compound group had to be described. The objective was to generate a description transferable to lactones with different ring size and also making it state transferable. Two parameter sets were produced, one reproducing only densities, and a second attempting to reproduce both densities and enthalpy of vaporization. The parameter sets were tested on three experimental data sets γ-butyrolactone - benzene at 293.15 K and δ-valerolactone - benzene at 293.15 K and 313.15 K. The predictions of the first parameter set gave score of the magnitude 10-20 and FST analysis of simulation using the second parameter set resulted in the prediction of liquid-liquid splits. The simple approach for generation of force field description has shown to be too simple when working with novel compound groups. However, the method is applicable in cases where parameters exist for a compound class, but adjustments are needed to reproduce a new compound of the same compound class.

The procedure for setting up, running and analyzing the MD simulations is simple and straightforward which is an advantage when compared to Gibbs Ensemble Monte Carlo or similar simulation methods where insertion moves are needed. Predictions by GEMC methods are in many cases not applicable to initial design decisions because errors in the prediction of pure component vapor pressures propagate into the mixture region. This problem is not encountered with the presented methodology provided experimental vapor pressures are known. It is questionable if the simulation of the vapor phase is needed at ambient pressures where the non-idealty if the vapor phase is insignificant. The work is based on open source software which reduces the cost compared to commercial software like COSMO*therm*. The methodology presented enables different approaches for handling non-ideal behavior in vapor phase opposite to COSMO*therm*. The methodology has the potential to become an important tool for chemical engineers in R&D groups due to its simplicity, robustness and

low cost.

A

Fluctuation Solution Theory

The theory is known as the statistical mechanical theory of solutions (Kirkwood and Buff, 1951). The main work lies in deriving the connection between the fluctuation in composition of the Grand Canonical Ensemble, GCE, and the radial distribution function, $g_{\alpha\beta}(R)$. Then they combine it with the relation between composition fluctuations and the chemical potential, μ, for the GCE and thereby get the relation between the radial distribution function and the derivative chemical potential with respect to composition. This part is divided into 4 parts: An introduction to the grand canonical ensemble, the derivation of the radial distribution function from composition fluctuations, the relation between the derivative of the chemical potential with respect to composition and the fluctuation in the composition, and last the relation between the radial distribution function and derivative of the chemical potential is shown.

A.1 Grand Canonical Ensemble

The grand canonical ensemble is also known as vTμ ensemble because the volume, the temperature and the chemical potential is held constant. The general form of the partition function, Ξ, is shown in Equation A.1.

$$\Xi = \sum_{N,i} \exp\left(\frac{N\mu - E_i}{kT}\right) \tag{A.1}$$

N is the number of molecules, E_i is the energy, k is the Boltzmann constant and T is the temperature. The average number of molecules, $\langle N \rangle_{Avg}$, in the simulation box is given by Equation A.2.

$$\langle N \rangle_{Avg} = \frac{\sum_{N,i} N \exp\left(\frac{N\mu - E_i}{kT}\right)}{\Xi} \tag{A.2}$$

Where the variation is given by:

$$\left\langle N^2 \right\rangle_{Avg} - \left(\langle N \rangle_{Avg}\right)^2 = kT \left(\frac{\partial \langle N \rangle_{Avg}}{\partial \mu}\right)_{v,T} \tag{A.3}$$

For multicomponent system the partition function is given by Equation A.4 and the average number molecules of type α is given by Equation A.5.

$$\Xi = \sum_{N_\alpha}\sum_{N_\beta}\cdots\sum_{i}\exp\left(\frac{N_\alpha\nu_\alpha + N_\beta\nu_\beta + \ldots - E_i}{kT}\right) \tag{A.4}$$

$$\langle N_\alpha\rangle = \frac{\sum_{N_\alpha}\sum_{N_\beta}\cdots\sum_i N_\alpha \exp\left(\frac{N_\alpha\nu_\alpha+N_\beta\nu_\beta+\ldots-E_i}{kT}\right)}{\Xi} \tag{A.5}$$

The general formulation in Equation A.3 can be extrapolated to multi-component systems where it becomes a matrix which elements take the form shown in Equation A.6.

$$\langle N_\alpha N_\beta\rangle_{Avg} - \langle N_\alpha\rangle_{Avg}\langle N_\beta\rangle_{Avg} = kT\left(\frac{\partial\langle N_\alpha\rangle_{Avg}}{\partial\mu_\beta}\right)_{v,T,N_\gamma} \tag{A.6}$$

where the subscript N_γ denotes that all other number densities are held constant.

The formulation shown above will be used in Section A.3 to derive the connection to the radial distribution function and derivative of the activity coefficient with respect to composition.

A.2 Radial Distribution Function

Before introducing the radial distribution function it is necessary to introduce the relations between average densities in systems of molecules and the fluctuation in density. One has to consider a volume, v, which is a part of an infinite big system. A test example of the grand ensemble which represents the behavior of the fluid (a.k.a. its statistical behavior). The volume contains N_1 molecules of type 1, N_2 of type 2,,, and N_ν of type ν. For specified configuration, $\mathbf{R}_{i_\alpha}$, of all the molecules the singlet density of component α is named $\nu_\alpha^{(1)}(\mathbf{R}_1)$ at point $\mathbf{R}_1$ in v and it is defined in Equation A.7. The density of ordered pairs of molecules of component α[1] at a point $\mathbf{R}_1$ and component β at a point $\mathbf{R}_2$ is named $\nu_{\alpha\beta}^{(2)}(\mathbf{R}_1,\mathbf{R}_2)$ and its definition is shown in equation A.8.

$$\nu_\alpha^{(1)}(\mathbf{R}_1) = \sum_{i_\alpha=1}^{N_\alpha}\delta\left(\mathbf{R}_{i_\alpha} - \mathbf{R}_1\right) \tag{A.7}$$

$$\nu_{\alpha\beta}^{(2)}(\mathbf{R}_1,\mathbf{R}_2) = \sum_{i_\alpha=1}^{N_\alpha}\sum_{k_\beta=1\,i_\alpha\neq k_\alpha}^{N_\beta}\delta\left(\mathbf{R}_{i_\alpha} - \mathbf{R}_1\right)\delta\left(\mathbf{R}_{k_\beta} - \mathbf{R}_2\right) \tag{A.8}$$

where $\delta\left(\mathbf{R}_{i_\alpha} - \mathbf{R}_1\right)$ is the 3-dimensional Dirac delta-function which is generally known as the impulse function Dirac (1958). The integrals over the space, v,

[1] Error in original article at this point. Where it says $\alpha\beta$ instead of only α.

of the two density functions are shown in Equations A.9 and A.10 which gives the nice results as shown.

$$\int^{v} \nu_{\alpha}^{(1)}(\mathbf{R}_1)dv_1 = N_\alpha \tag{A.9}$$

$$\int^{v} \nu_{\alpha\beta}^{(2)}(\mathbf{R}_1,\mathbf{R}_2)dv_1dv_2 = N_\alpha N_\beta - N_\alpha\delta_{\alpha\beta} \tag{A.10}$$

where $\delta_{\alpha\beta}$ in Equation A.10 is the Kronecker delta[2].

Now define the same properties for the grand canonical ensemble and they are called average number densities which is the result of the fluctuations in the number of molecules. The equivalent of Equations A.7 and A.8 are shown in Equations A.11 and A.12.

$$\rho_{\alpha}^{(1)}(\mathbf{R}_1) = \left\langle \nu_{\alpha}^{(1)} \right\rangle_{Avg} \tag{A.11}$$

$$\rho_{\alpha\beta}^{(2)}(\mathbf{R}_1,\mathbf{R}_2) = \left\langle \nu_{\alpha\beta}^{(2)} \right\rangle_{Avg} \tag{A.12}$$

The equations above can be integrated to find the mean number densities of the volume, v, both for the singlet density and the density of ordered pairs.

$$\int^{v} \rho_{\alpha}^{(1)}(\mathbf{R}_1)dv_1 = \langle N_\alpha \rangle_{Avg} \tag{A.13}$$

$$\int^{v}\int^{v} \rho_{\alpha\beta}^{(2)}(\mathbf{R}_1,\mathbf{R}_2)dv_1dv_2 = \left\langle N_\alpha N_\beta^{(2)} \right\rangle_{Avg} - \delta_{\alpha\beta}\langle N_\alpha \rangle_{Avg} \tag{A.14}$$

To introduce the radial distribution function, $g_{\alpha\beta}^{(2)}(\mathbf{R})$ we make an integral of the difference between the density of ordered pairs and the product of their singlet densities.

$$\int^{v}\int^{v} \rho_{\alpha\beta}^{(2)}(\mathbf{R}_1,\mathbf{R}_2) - \rho_{\alpha}^{(1)}(\mathbf{R}_1)\rho_{\beta}^{(1)}(\mathbf{R}_2)dv_1dv_2$$

$$= \left[\left\langle N_\alpha N_\beta^{(2)} \right\rangle_{Avg} - \langle N_\alpha \rangle_{Avg}\langle N_\beta \rangle_{Avg}\right] - \delta_{\alpha\beta}\langle N_\alpha \rangle_{Avg} \tag{A.15}$$

The integral in Equation A.15 can also be calculated using the mean densities which are shown in the equations below. These are mean densities and can be used for all fluids, both liquids and gases.

$$\rho_{\alpha}^{(1)}(\mathbf{R}_1) = \frac{\langle N_\alpha \rangle_{Avg}}{v} = c_\alpha \tag{A.16}$$

$$\rho_{\alpha\beta}^{(2)}(\mathbf{R}_1,\mathbf{R}_2) = c_\alpha c_\beta g_{\alpha\beta}^{(2)}(\mathbf{R}) \tag{A.17}$$

$$\mathbf{R} = (\mathbf{R}_1 - \mathbf{R}_2)$$

[2] $\delta_{\alpha\beta} = 1$ for $\alpha = \beta$ and $\delta_{\alpha\beta} = 0$ for $\alpha \neq \beta$

Below Equations A.16 and A.17 are introduce into the integral in Equation A.15.

$$\begin{aligned}\int^v \int^v & \rho^{(2)}_{\alpha\beta}(\mathbf{R}_1,\mathbf{R}_2) - \rho^{(1)}_{\alpha}(\mathbf{R}_1)\rho^{(1)}_{\beta}(\mathbf{R}_2)dv_1dv_2 \\ &= \int^v \int^v c_\alpha c_\beta g^{(2)}_{\alpha\beta}(\mathbf{R}) - c_\alpha c_\beta dv_1 dv_2 \\ &= c_\alpha c_\beta v \int^v \left(g^{(2)}_{\alpha\beta}(\mathbf{R}) - 1\right) dv \\ &= \frac{\langle N_\alpha \rangle_{Avg} \langle N_\beta \rangle_{Avg}}{v} \int^v \left(g^{(2)}_{\alpha\beta}(\mathbf{R}) - 1\right) dv \end{aligned} \tag{A.18}$$

Now it is possible to combine Equations A.15 and A.18 to introduce the connection between the radial distribution function and the fluctuations in the composition.

$$\int^v \left(g^{(2)}_{\alpha\beta}(\mathbf{R}) - 1\right) dv = v \frac{\left\langle N_\alpha N^{(2)}_\beta \right\rangle_{Avg} - \langle N_\alpha \rangle_{Avg} \langle N_\beta \rangle_{Avg}}{\langle N_\alpha \rangle_{Avg} \langle N_\beta \rangle_{Avg}} - \frac{\delta_{\alpha\beta}}{c_\alpha} \tag{A.19}$$

In the last term on the right hand side the index should be β but the term is only included when $\alpha = \beta$ thereby it have no influence whether the index is α or β. The original article uses α and to assure consistency α is also used here.

A.3 Chemical Potential Relation

In Section A.1 Equation A.6 showed the relation between the fluctuation in number density and the derivative of the average number density with respect to the chemical potential at constant temperature and volume. In this section the objective is to find the expression with the derivative of the chemical potential with respect to the composition at constant temperature and pressure. The shift to pressure dependency instead of volume dependency is made because the Gibbs-Duhem equation (Equation A.27) has to be applied later.

First we define two matrices $\mathbf{B}$ which contains the elements generated using Equation A.6 and $\mathbf{G}$ which contains the integral over volume of the radial distribution function.

$$\mathbf{B}_{\alpha\beta} = kT \left(\frac{\partial N_\alpha}{\partial \mu_\beta}\right)_{v,T,N_\gamma} \tag{A.20}$$

$$\mathbf{G}_{\alpha\beta} = \int^v \left(g^{(2)}_{\alpha\beta}(\mathbf{R}) - 1\right) dv \tag{A.21}$$

By combination of Equations A.6, A.16, A.19, A.20 and A.21 the elements of $\mathbf{B}$ can be expressed by the elements of $\mathbf{G}$.

$$\mathbf{B} = c_\alpha c_\beta \mathbf{G}_{\alpha\beta} + c_\alpha \delta_{\alpha\beta} \tag{A.22}$$

To get closer to the goal we get the inverse of the **B** matrix to find the derivative of the chemical potential, $d\mu_\alpha/dN_\beta$ as shown in Equation A.23.

$$\frac{1}{kT}\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,V,N_\gamma} = \frac{|\mathbf{B}|_{\alpha\beta}}{v|\mathbf{B}|} \tag{A.23}$$

$|B|_{\alpha\beta}$ is the cofactor of element $\mathbf{B}_{\alpha\beta}$ and $|\mathbf{B}|$ is the determinant of the matrix. They are explained further in Appendix B.1. As mentioned above we want to shift dependency from volume to pressure which is done using Equation A.24 which derived in Appendix B.2.

$$\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,V,N_\gamma} = \left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\gamma} + \frac{\bar{v}_\alpha \bar{v}_\beta}{\kappa v} \tag{A.24}$$

where $\bar{v}_i$ is the partial molar volume of component i and κ is the compressibility of the mixture. By introducing Equation A.23 into Equation A.24 it is possible to get a preliminary result where $(\partial\mu_\alpha/\partial N_\beta)_{TPN_\gamma}$ is described by the elements of the **B**-matrix.

$$\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\gamma} = \frac{kT|\mathbf{B}|_{\alpha\beta}}{v|\mathbf{B}|} - \frac{\bar{v}_\alpha \bar{v}_\beta}{\kappa v} \tag{A.25}$$

Using Equations A.26 and A.27 together with Equation A.25 it is possible to derive the general Equations A.28-A.30. The mathematical operations to get the resulting equations are available in Appendix B.3.

$$1 = \sum_{\alpha=1}^{v} c_\alpha \bar{v}_\alpha \tag{A.26}$$

$$0 = \sum_{\alpha=1}^{v} N_\alpha \left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\gamma} \tag{A.27}$$

$$kT\kappa = \frac{|\mathbf{B}|}{\sum_{\alpha=1}^{v}\sum_{\beta=1}^{v} c_\alpha c_\beta |\mathbf{B}|_{\alpha\beta}} \tag{A.28}$$

$$\bar{v}_\alpha = \frac{\sum_{\omega=1}^{v} c_\omega |\mathbf{B}|_{\alpha\omega}}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}} \tag{A.29}$$

$$\frac{v}{kT}\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\gamma} = \left(\frac{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma \left(\mathbf{B}^{-1}_{\beta\alpha}\mathbf{B}^{-1}_{\gamma\omega} - \mathbf{B}^{-1}_{\omega\alpha}\mathbf{B}^{-1}_{\gamma\beta}\right)}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma \mathbf{B}^{-1}_{\gamma\omega}}\right) \tag{A.30}$$

Where $\mathbf{B}^{-1}_{ij}$ is the ij-element of the inverse **B**-matrix. The formulation in Equations A.28, A.29 and A.30 can seem very complex. To simplify the resulting relations of the binary case is shown below[3]:

$$kT\kappa = \frac{1 + c_1 G_{11} + c_2 G_{22} + c_1 c_2 \left(G_{11}G_{22} - G_{12}^2\right)}{c_1 + c_2 + c_1 c_2 \left(G_{11} + G_{22} - 2G_{12}\right)} \tag{A.31}$$

[3]There is an error in the original article by Kirkwood and Buff (1951). In the nominator of the expression for $kT\kappa$ the brackets are given like: $(G_{11}G_{22} - G_{12})^2$ where the squaring has been misplaced.

$$\bar{v}_1 = \frac{1 + c_2 (G_{22} - G_{12})}{c_1 + c_2 + c_1 c_2 (G_{11} + G_{22} - 2G_{12})} \tag{A.32}$$

$$\frac{v}{kT} \left(\frac{\partial \mu_1}{\partial N_1} \right)_{T,P,N_2} = \frac{c_2}{\rho c_1 + c_1^2 c_2 (G_{11} + G_{22} - 2G_{12})} \tag{A.33}$$

The mathematically procedure to get to these results are shown in Appendix B.4. Last mole number densities are replaced by mole fraction densities in the derivative.

$$\begin{aligned}
N \left(\frac{\partial \mu_1}{\partial N_1} \right)_{T,P,N_2} &= N \left(\frac{\partial \mu_1}{\partial x_1} \right)_{T,P} \left(\frac{\partial x_1}{\partial N_1} \right)_{T,P,N_2} \\
&= N \left(\frac{\partial \mu_1}{\partial x_1} \right)_{T,P} \frac{1 - x_1}{N} \Leftrightarrow \\
\left(\frac{\partial \mu_1}{\partial x_1} \right)_{T,P} &= \frac{N}{1 - x_1} \left(\frac{\partial \mu_1}{\partial N_1} \right)_{T,P,N_2} \\
&= \frac{N}{1 - x_1} \frac{kT}{v} \frac{c_2}{\rho c_1 + c_1^2 c_2 (G_{11} + G_{22} - 2G_{12})} \\
&= kT \frac{1}{x_1} \frac{1}{1 + x_1 c_2 (G_{11} + G_{22} - 2G_{12})} \\
&= kT \left(\frac{1}{x_1} - \frac{c_2 (G_{11} + G_{22} - 2G_{12})}{1 + x_1 c_2 (G_{11} + G_{22} - 2G_{12})} \right)
\end{aligned} \tag{A.34}$$

With the equations available for linking between the integrals of the radial distribution function and the derivative of the chemical potential only a single step is missing before it theory is applicable to activity coefficient models.

A.4 Advances by O'Connell

O'Connell (1971a,b) continued the theoretical work and derived the expression (Eq. A.35) which is the foundation of the methodology of this thesis:

$$\begin{aligned}
\frac{1}{kT} \left(\frac{\partial \mu_1}{\partial x_1} \right)_{T,P} &= \frac{1}{kT} \left(\frac{\partial}{\partial x_1} \right)_{T,P} (\ln \gamma_1 + \ln x_1) \\
&= \frac{1}{kT} \left(\left(\frac{\partial \ln \gamma_1}{\partial x_1} \right)_{T,P} + \left(\frac{\partial \ln x_1}{\partial x_1} \right)_{T,P} \right) \Leftrightarrow \\
\left(\frac{\partial \ln \gamma_1}{\partial x_1} \right)_{T,P} &= \left(\frac{\partial \mu_1}{\partial x_1} \right)_{T,P} - \frac{1}{x_1} \\
&= -\frac{c_2 (G_{11} + G_{22} - 2G_{12})}{1 + x_1 c_2 (G_{11} + G_{22} - 2G_{12})} \\
&= -\frac{x_2 (H_{11} + H_{22} - 2H_{12})}{1 + x_1 x_2 (H_{11} + H_{22} - 2H_{12})}
\end{aligned} \tag{A.35}$$

In Equation A.35 the total correlation function integral, TCFI or H_{ij}, is given by G_{ij}/v. In the TCFI's are preferred as they are unit less. Similar expressions can be derived for isothermal compressibility and the partial molar volume.

$$\rho\kappa_T kT = \frac{1 + x_1 H_{11} + x_2 H_{22} + x_1 x_2 \left(H_{11}H_{22} - H_{12}^2\right)}{1 + x_1 x_2 \Delta H} \tag{A.36}$$

$$\rho\bar{V}_1 = \frac{1 + x_2 \left(H_{22} - H_{12}\right)}{1 + x_1 x_2 \Delta H} \tag{A.37}$$

Expressions can also be derived for ternary mixtures where the expressions get comprehensive (O'Connell, 1971b).

B

FST Supplements

In the sections below you will find elaborated derivations of certain equations which were left out of the main text.

B.1 Inversion of a Matrix

Here it is shown how to generate the inverse of a 2×2 and a 3×3 matrix. First the smallest matrix of the two, where $\mathbf{I}$ is the unit matrix:

$$\mathbf{I} = \mathbf{A}\mathbf{A}^{-1} \tag{B.1}$$

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{B.2}$$

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{B.3}$$

$$= \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \tag{B.4}$$

The same procedure can be made for the 3×3 matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{B.5}$$

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{13} & a_{12} \\ a_{33} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ \begin{vmatrix} a_{23} & a_{21} \\ a_{33} & a_{31} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} & \begin{vmatrix} a_{12} & a_{11} \\ a_{32} & a_{31} \end{vmatrix} & \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \end{bmatrix} \tag{B.6}$$

$$\tag{B.7}$$

Each element in the inverse matrix neglecting the prefactor $1/|\mathbf{A}|$ is also called the co-factor of that element in the $\mathbf{A}$-matrix and the symbo normally used is $|\mathbf{A}_{\alpha\beta}|$. Thereby the each element in the inverse matrix can be symbolized by $|\mathbf{A}_{\alpha\beta}|/|\mathbf{A}|$

The procedure is more complicated for bigger dimensions and will not be shown here but the symbolism with respect to the co-factor can also be used for bigger matrices.

B.2 Switch dependency of μ_i from constant V to P

The change in the chemical potential of compound i is given below

$$d\mu_i|_{T,P,n_{j\neq i}} = \left(\frac{\partial \mu_i}{\partial T}\right)_{P,n} dT + \left(\frac{\partial \mu_i}{\partial P}\right)_{T,n} dP + \sum_j \left(\frac{\partial \mu_i}{\partial n_j}\right)_{T,P,n_{k\neq j}} dn_j \tag{B.8}$$

The equation can extracted by replacing dP with the expression:

$$dP = \left(\frac{\partial P}{\partial V}\right)_{T,n} dV + \left(\frac{\partial P}{\partial T}\right)_{V,n} \partial T + \sum_j \left(\frac{\partial P}{\partial n_j}\right)_{T,V,n_{k\neq j}} dn_j \tag{B.9}$$

If equation B.9 is inserted into B.8 then dP can be eliminated and it is possible to get an expression dependent of the variables dT, dV and dn_j.

$$\begin{aligned} d\mu_i|_{T,P,n_{j\neq i}} = & \left[\left(\frac{\partial \mu_i}{\partial T}\right)_{P,n} + \left(\frac{\partial \mu_i}{\partial P}\right)_{T,n}\left(\frac{\partial P}{\partial T}\right)_{V,n}\right] dT \\ & + \left[\left(\frac{\partial \mu_i}{\partial P}\right)_{T,n}\left(\frac{\partial P}{\partial V}\right)_{T,n}\right] dV \\ & + \sum_j \left[\left(\frac{\partial \mu_i}{\partial n_j}\right)_{T,P,n_{k\neq j}} + \left(\frac{\partial \mu_i}{\partial P}\right)_{T,n}\left(\frac{\partial P}{\partial n_j}\right)_{T,V,n_{k\neq j}}\right] dn_j \end{aligned} \tag{B.10}$$

Then equation B.10 can be derived with respect to ∂n_j at constant temperature, volume and mole numbers of other compounds

$$\begin{aligned} \left(\frac{\partial \mu_i}{\partial n_j}\right)_{T,V,n_{k\neq j}} &= \left(\frac{\partial \mu_i}{\partial n_j}\right)_{T,P,n_{k\neq j}} + \left(\frac{\partial \mu_i}{\partial P}\right)_{T,n}\left(\frac{\partial P}{\partial n_j}\right)_{T,V,n_{k\neq j}} \\ &= \left(\frac{\partial \mu_i}{\partial n_j}\right)_{T,P,n_{k\neq j}} - \bar{V}_i \left(\frac{\partial V}{\partial n_i}\right)_{T,P,n_{i\neq j}} \left(\frac{\partial P}{\partial V}\right)_{T,\mathbf{n}} \\ &= \left(\frac{\partial \mu_i}{\partial n_j}\right)_{T,P,n_{k\neq j}} + \frac{\bar{V}_i \bar{V}_j}{\kappa V} \end{aligned} \tag{B.11}$$

In the derivation above the following relations have been used

$$\kappa = -\frac{1}{V}\left(\frac{\partial V}{\partial P}\right)_{T,\mathbf{n}} \tag{B.12}$$

$$\bar{V}_i = \left(\frac{\partial V}{\partial n_i}\right)_{T,P,n_{j\neq i}} \tag{B.13}$$

$$= -\left(\frac{\partial P}{\partial n_i}\right)_{T,V,n_{j\neq i}} / \left(\frac{\partial P}{\partial V}\right)_{T,\mathbf{n}} \tag{B.14}$$

$$= \left(\frac{\partial \mu_i}{\partial P}\right)_{T,\mathbf{n}} \tag{B.15}$$

B.3 Derivation of properties general case

But to get the final expression it is needed to express the partial volumes and the compressibility factor by elements of the **B**-matrix. First the compressibility factor is determined by applying the Gibbs-Duhem equation (Eq. B.17).

$$\sum_{\alpha=1}^{v} c_\alpha \bar{v}_\alpha = 1 \tag{B.16}$$

$$\sum_{\alpha=1}^{v} N_\alpha \left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\gamma} = 0 \tag{B.17}$$

$$= \sum_{\alpha=1}^{v} \frac{N_\alpha}{v}\left(kT\frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} - \frac{\bar{v}_\alpha \bar{v}_\beta}{\kappa}\right)$$

$$= \sum_{\alpha=1}^{v} kT c_\alpha \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} - \sum_{\alpha=1}^{v} c_\alpha \frac{\bar{v}_\alpha \bar{v}_\beta}{\kappa}$$

$$= \sum_{\alpha=1}^{v} kT c_\alpha \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} - \frac{\bar{v}_\beta}{\kappa}\sum_{\alpha=1}^{v} c_\alpha \bar{v}_\alpha \Leftrightarrow$$

$$\frac{\bar{v}_\beta}{\kappa} = kT \sum_{\alpha=1}^{v} c_\alpha \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} \tag{B.18}$$

The Gibbs-Duhem can also be formulated in another way (Equation B.19) as shown below which gives a variation of Equation B.18.

$$\sum_{\beta=1}^{v} N_\beta \left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\gamma} = 0 \tag{B.19}$$

$$= \sum_{\beta=1}^{v} \frac{N_\beta}{v}\left(kT\frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} - \frac{\bar{v}_\alpha \bar{v}_\beta}{\kappa}\right)$$

$$= \sum_{\beta=1}^{v} kT c_\beta \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} - \sum_{\beta=1}^{v} c_\beta \frac{\bar{v}_\alpha \bar{v}_\beta}{\kappa}$$

$$= \sum_{\beta=1}^{v} kT c_\beta \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} - \frac{\bar{v}_\alpha}{\kappa}\sum_{\beta=1}^{v} c_\beta \bar{v}_\beta \Leftrightarrow$$

$$\frac{\bar{v}_\alpha}{\kappa} = kT\sum_{\beta=1}^{v} c_\beta \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} \tag{B.20}$$

The formulation shown in Equation B.18 is used by Kirkwood and Buff in their article Kirkwood and Buff (1951) and will be used here. The second last step in the isolation of κ is done using Equations A.26 and B.18 where the c_β is multiplied onto the eqautions.

$$\begin{aligned}
\sum_{\alpha=1}^{v} \frac{c_\alpha \bar{v}_\alpha}{\kappa} &= \sum_{\alpha=1}^{v} c_\alpha \sum_{\beta=1}^{v} kT c_\beta \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} \Leftrightarrow \\
\frac{1}{\kappa} &= \frac{kT}{|\mathbf{B}|}\sum_{\alpha=1}^{v} c_\alpha \sum_{\beta=1}^{v} c_\beta |\mathbf{B}|_{\alpha\beta} \Leftrightarrow \\
kT\kappa &= \frac{|\mathbf{B}|}{\sum_{\alpha=1}^{v} c_\alpha \sum_{\beta=1}^{v} c_\beta |\mathbf{B}|_{\alpha\beta}} \\
&= \frac{|\mathbf{B}|}{\sum_{\alpha=1}^{v}\sum_{\beta=1}^{v} c_\alpha c_\beta |\mathbf{B}|_{\alpha\beta}}
\end{aligned} \tag{B.21}$$

Equation B.21 can be inserted into Equation B.20 and a general expression for the partial molar volume is the result.

$$\bar{v}_\alpha = \frac{\sum_{\omega=1}^{v} c_\omega |\mathbf{B}|_{\alpha\omega}}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}} \tag{B.22}$$

with all properties in Equation A.25 described by the **B**-matrix it is possible to derive a general form the later equation only using elements from **B**.

$$\begin{aligned}
\frac{v}{kT}\left(\frac{\partial \mu_\alpha}{\partial N_\beta}\right)_{T,P,N_\gamma} &= \frac{|\mathbf{B}|_{\alpha\beta}}{|\mathbf{B}|} - \frac{\frac{\sum_{\omega=1}^{v} c_\omega |\mathbf{B}|_{\alpha\omega} \sum_{\omega=1}^{v} c_\omega |\mathbf{B}|_{\beta\omega}}{\left(\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}\right)^2}}{\frac{|\mathbf{B}|}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}}} \\
&= \frac{1}{|\mathbf{B}|}\left(|\mathbf{B}|_{\alpha\beta} - \frac{\sum_{\omega=1}^{v} c_\omega |\mathbf{B}|_{\alpha\omega} \sum_{\gamma=1}^{v} c_\gamma |\mathbf{B}|_{\beta\gamma}}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}}\right) \\
&= \frac{1}{|\mathbf{B}|}\left(\frac{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\alpha\beta}|\mathbf{B}|_{\omega\gamma}}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}}\right. \\
&\quad \left. - \frac{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\alpha\omega}|\mathbf{B}|_{\beta\gamma}}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}}\right) \\
&= \frac{1}{|\mathbf{B}|}\left(\frac{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma \left(|\mathbf{B}|_{\alpha\beta}|\mathbf{B}|_{\omega\gamma} - |\mathbf{B}|_{\alpha\omega}|\mathbf{B}|_{\beta\gamma}\right)}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma |\mathbf{B}|_{\omega\gamma}}\right) \\
&= \left(\frac{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma \left(\mathbf{B}^{-1}_{\beta\alpha}\mathbf{B}^{-1}_{\gamma\omega} - \mathbf{B}^{-1}_{\omega\alpha}\mathbf{B}^{-1}_{\gamma\beta}\right)}{\sum_{\omega=1}^{v}\sum_{\gamma=1}^{v} c_\omega c_\gamma \mathbf{B}^{-1}_{\gamma\omega}}\right)
\end{aligned} \tag{B.23}$$

B.4 Derivation of properties for binary case

Here the mathematically steps from the general formulations of to the final expressions are shown for the binary system. The first derivation is for the compressibility factor, κ. First the $\mathbf{B}$-matrix is shown.

$$\begin{aligned}
\mathbf{B} &= \begin{bmatrix} c_1 + c_1^2 G_{11} & c_1 c_2 G_{12} \\ c_1 c_2 G_{12} & c_2 + c_2^2 G_{22} \end{bmatrix} \\
kT\kappa &= \frac{|\mathbf{B}|}{c_1^2|B_{11}| + c_2^2|B_{22}| + 2c_1c_2|B_{12}|} \\
&= \frac{\left(c_1 + c_1^2 G_{11}\right)\left(c_2 + c_2^2 G_{22}\right) - c_1^2 c_2^2 G_{12}}{c_1^2\left(c_2 + c_2^2 G_{22}\right) + c_2^2\left(c_1 + c_1^2 G_{11}\right) - 2c_1^2 c_2^2 G_{12}} \\
&= \frac{c_1c_2 + c_1c_2^2 G_{22} + c_1^2 c_2 G_{11} + c_1^2 c_2^2 G_{11}G_{22} - c_1^2 c_2^2 G_{12}}{c_1^2 c_2 + c_1^2 c_2^2 G_{22} + c_1 c_2^2 + c_1^2 c_2^2 G_{11} - 2c_1^2 c_2^2 G_{12}} \\
&= \frac{1 + c_1 G_{11} + c_2 G_{22} + c_1 c_2\left(G_{11}G_{22} - G_{12}\right)}{c_1 + c_2 + c_1 c_2 G_{11} + c_1 c_2 G_{22} - 2c_1 c_2 G_{12}} \\
&= \frac{1 + c_1 G_{11} + c_2 G_{22} + c_1 c_2\left(G_{11}G_{22} - G_{12}\right)}{c_1 + c_2 + c_1 c_2\left(G_{11} + G_{22} - 2G_{12}\right)}
\end{aligned}$$

Next is the partial molar volume, $\bar{v}_1$.

$$\begin{aligned}
\bar{v}_1 &= \frac{c_1|\mathbf{B}_{11}| + c_2|\mathbf{B}_{12}|}{c_1^2|\mathbf{B}_{11}| + c_2^2|\mathbf{B}_{22}| + 2c_1c_2|\mathbf{B}_{12}|} \\
&= \frac{c_1\left(c_2 + c_2^2 G_{22}\right) - c_1 c_2^2 G_{12}}{c_1^2\left(c_2 + c_2^2 G_{22}\right) + c_2^2\left(c_1 + c_1^2 G_{11}\right) - 2c_1^2 c_2^2 G_{12}} \\
&= \frac{1 + c_2 G_{22} - c_2 G_{12}}{c_1 + c_2 + c_2^2 + c_1 c_2\left(G_{11}G_{22} - 2G_{12}\right)}
\end{aligned}$$

Only one of the four possible derivatives is shown here.

$$\begin{aligned}
\frac{v}{kT}\left(\frac{\partial \mu_1}{\partial N_1}\right)_{T,P,N_2} &= \left(\frac{\sum_{\omega=1}^{2}\sum_{\gamma=1}^{2} c_\omega c_\gamma\left(\mathbf{B}_{11}^{-1}\mathbf{B}_{\gamma\omega}^{-1} - \mathbf{B}_{\omega 1}^{-1}\mathbf{B}_{\gamma 1}^{-1}\right)}{\sum_{\omega=1}^{2}\sum_{\gamma=1}^{2} c_\omega c_\gamma \mathbf{B}_{\gamma\omega}^{-1}}\right) \\
&= \frac{1}{c_1^2\mathbf{B}_{11}^{-1} + c_2^2\mathbf{B}_{22}^{-1} + 2c_1c_2\mathbf{B}_{12}^{-1}} \\
&\quad \left(c_1^2\left(\mathbf{B}_{11}^{-1}\mathbf{B}_{11}^{-1} - \mathbf{B}_{11}^{-1}\mathbf{B}_{11}^{-1}\right)\right. \\
&\quad + c_1c_2\left(\mathbf{B}_{11}^{-1}\mathbf{B}_{12}^{-1} - \mathbf{B}_{11}^{-1}\mathbf{B}_{21}^{-1}\right) \\
&\quad + c_2c_1\left(\mathbf{B}_{11}^{-1}\mathbf{B}_{21}^{-1} - \mathbf{B}_{21}^{-1}\mathbf{B}_{11}^{-1}\right) \\
&\quad \left. + c_2^2\left(\mathbf{B}_{11}^{-1}\mathbf{B}_{22}^{-1} - \mathbf{B}_{21}^{-1}\mathbf{B}_{21}^{-1}\right)\right) \\
&= \frac{c_2^2\left(\mathbf{B}_{11}^{-1}\mathbf{B}_{22}^{-1} - \mathbf{B}_{21}^{-1}\mathbf{B}_{21}^{-1}\right)}{c_1^2\mathbf{B}_{11}^{-1} + c_2^2\mathbf{B}_{22}^{-1} + 2c_1c_2\mathbf{B}_{12}^{-1}} \\
&= \frac{c_2^2|\mathbf{B}|}{c_1^2\mathbf{B}_{11}^{-1} + c_2^2\mathbf{B}_{22}^{-1} + 2c_1c_2\mathbf{B}_{12}^{-1}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{|\mathbf{B}|}\frac{c_2^2|\mathbf{B}|}{c_1^2\mathbf{B}_{22}+c_2^2\mathbf{B}_{11}-2c_1c_2\mathbf{B}_{12}} \\
&= \frac{c_2^2}{c_2^2\mathbf{B}_{11}+c_1^2\mathbf{B}_{22}-2c_1c_2\mathbf{B}_{12}} \\
&= \frac{c_2^2}{c_2^2\left(c_1+c_1^2G_{11}\right)+c_1^2\left(c_2+c_2^2G_{22}\right)-2c_1^2c_2^2G_{12}} \\
&= \frac{c_2}{c_2\left(c_1+c_1^2G_{11}\right)+c_1^2\left(1+c_2G_{22}\right)-2c_1^2c_2G_{12}} \\
&= \frac{c_2}{c_2c_1+c_1^2+c_1^2c_2\left(G_{11}+G_{22}-2G_{12}\right)} \\
&= \frac{c_2}{\rho c_1+c_1^2c_2\left(G_{11}+G_{22}-2G_{12}\right)} \qquad \text{(B.24)} \\
&= \frac{1}{c_1}\frac{c_2}{c_1+c_2+c_1c_2\left(G_{11}+G_{22}-2G_{12}\right)} \qquad \text{(B.25)}
\end{aligned}
$$

where ρ is the number density as given in Equation B.26 for the binary case.

$$\rho = \frac{N}{v} = \frac{N_1}{v} + \frac{N_2}{v} = c_1 + c_2 \qquad \text{(B.26)}$$

B.5 From Mole Numbers to Mole Fractions

For a binary composition it is to change from mole numbers to mole fractions.

$$
\begin{aligned}
\frac{1}{kT}\left(\frac{\partial\mu_1}{\partial x_1}\right) &= \frac{1}{kT}\frac{N}{1-x_1}\left(\frac{\partial\mu_1}{\partial N_1}\right)_{T,P,N_2} \\
&= \frac{N}{v\left(1-x_1\right)}\frac{1}{c_1}\frac{c_2}{c_1+c_2+c_1c_2\left(G_{11}+G_{22}-2G_{12}\right)} \\
&= \frac{N}{vx_2}\frac{x_2}{x_1}\frac{1}{\frac{N}{v}+\frac{N}{v}x_1c_2\left(G_{11}+G_{22}-2G_{12}\right)} \\
&= \frac{1}{x_1}\frac{1}{1+x_1c_2\left(G_{11}+G_{22}-2G_{12}\right)} \\
&= \frac{1}{x_1}-\frac{c_2\left(G_{11}+G_{22}-2G_{12}\right)}{1+x_1c_2\left(G_{11}+G_{22}-2G_{12}\right)} \qquad \text{(B.27)}
\end{aligned}
$$

where Equation B.28 has been used to reduce the expression.

$$\frac{c_2}{c_1} = \frac{\frac{N_2}{v}\frac{1}{N}}{\frac{N_1}{v}\frac{1}{N}} = \frac{x_2}{x_1} \qquad \text{(B.28)}$$

C

CHARMM Parameter Tables

atom	description	q
benzene		
CA	aromatic carbon	-0.115
HP	aromatic hydrogen	0.115
ethanol		
H	alcohol hydrogen	0.430
OH1	alcohol oxygen	-0.660
CT2	methylene carbon	0.050
CT3	methyl carbon	-0.270
HA	alkane hydrogen	0.09
methyl acetate		
CT3	methyl carbon (acetate)	-0.17
CD	acetate carbon	0.63
OB	acetate oxygen (double bond)	-0.52
OS	acetate oxygen (single bond)	-0.34
CT3	methyl carbon (single oxygen bond)	-0.14
HA	alkane hydrogen	0.09
n-pentane		
CT3	methyl carbon	-0.27
CT2	methylene carbon	-0.18
HA	alkane hydrogen	0.09
acetone		
CT3	methyl carbon	-0.27
HA	alkane hydrogen	0.09
CC	ketone carbon	0.55
O	ketone oxygen	-0.55

Table C.1. Atom descriptions and charges taken from the CHARMM force field.

atom	description	q
(2H)-heptafluoropropane		
CF1	single fluoro bonded carbon	0.11
CF3	triple fluoro bonded carbon	0.45
HF1	hydrogen	0.11
F1	fluor bonded to CF1	-0.22
F3	fluor bonded to CF3	-0.15
cyclohexane		
CC2	cyclic methylene carbon	-0.18
HA	alkane hydrogen	0.09
cyclopentane		
CC2	cyclic methylene carbon	-0.18
HA	alkane hydrogen	0.09
γ-butyrolactone		
CC2	ether bonded methylene carbon	-0.07
CC2	cyclic methylene carbon	-0.18
CC2	ester neighbour methylene carbon	-0.10
CDC	ester carbon	0.67
OBC	double bonded ester oxygen	-0.52
OSC	single bonded ester oxygen	-0.34
δ-valerolactone		
CC2	ether bonded methylene carbon	-0.07
CC2	cyclic methylene carbon	-0.18
CC2	ester neighbour methylene carbon	-0.10
CDC	ester carbon	0.67
OBC	double bonded ester oxygen	-0.52
OSC	single bonded ester oxygen	-0.34

Table C.2. Atom descriptions and charges taken from the CHARMM force field.

bond	K_b/kcal/mole/A^2	b_0/A
CA-CA	305.0	1.375
CA-HP	340.0	1.080
OH1-H	545.0	0.960
CT2-OH1	428.0	1.420
CT2-CT3	222.5	1.538
CT2-HA	309.0	1.111
CT3-HA	322.0	1.111
CT3-CD	200.0	1.522
OB-CD	750.0	1.220
OS-CD	150.0	1.334
OS-CT3	340.0	1.430
CT2-CT2	222.5	1.530
CT3-CT2	222.5	1.528
CT3-CC	200.0	1.522
O-CC	650.0	1.230
CF1-CF3	230.0	1.520
F1-CF1	420.0	1.374
F3-CF3	265.0	1.340
CF1-HF1	342.0	1.0828
CC2-CC2	222.5	1.530
CC2-HA	309.0	1.111
CC2-CDC	200.0	1.522
OSC-CC2	340.0	1.430
OBC-CDC	750.0	1.220
OSC-CDC	150.0	1.334

Table C.3. Bond vibration parameters taken from the CHARMM force field.

angle	K_θ $\frac{\text{kcal}}{\text{mole}\cdot\text{rad}^2}$	θ_0	K_{UB} $\frac{\text{kcal}}{\text{mole}\cdot\text{A}^2}$	S_0 A
CA-CA-CA	40.00	120.00	35.00	2.4162
CA-CA-HP	30.00	120.00	22.00	2.1525
H-OH1-CT2	57.50	106.00	···	···
OH1-CT2-CT3	75.70	110.10	···	···
OH1-CT2-HA	45.90	108.89	···	···
CT2-CT3-HA	34.60	110.10	22.53	2.1790
CT3-CT2-HA	34.60	110.10	22.53	2.1790
HA-CT2-HA	35.50	109.00	5.40	1.8020
HA-CT3-HA	35.50	108.40	5.40	1.8020
CT3-OS-CD	40.00	109.60	30.00	2.2651
HA-CT3-CD	33.00	109.50	30.00	2.1630
OB-CD-CT3	70.00	125.00	20.00	2.4420
OS-CD-CT3	55.00	109.00	20.00	2.3260
OS-CD-OB	90.00	125.90	160.00	2.2576
OS-CT3-HA	60.00	109.50	···	···
CT2-CT2-CT2	58.35	113.60	11.16	2.5610
CT2-CT2-CT3	58.00	115.00	8.00	2.5610
CT2-CT2-HA	26.50	110.10	22.53	2.1790
CT3-CC-CT3	50.00	116.50	50.00	2.4500
HA-CT3-CC	33.00	109.50	30.00	2.1630
O-CC-CT3	15.00	121.00	50.00	2.4400
F3-CF3-F3	118.00	107.00	30.00	2.1550
F1-CF1-HF1	57.50	108.89	5.00	1.9970
CF3-CF1-CF3	53.35	111.00	8.00	2.5610
HF1-CF1-CF3	34.50	110.10	22.53	2.1790
F3-CF3-CF1	42.00	112.00	30.00	2.3570
F1-CF1-CF3	44.00	112.00	30.00	2.3690
CC2-CC2-CC2	58.35	113.60	11.16	2.5610
HA-CC2-CC2	26.50	110.10	22.53	2.1790
CC2-CC2-CDC	52.00	108.00	···	···
CC2-OSC-CDC	40.00	109.60	30.00	2.2651
HA-CC2-CDC	33.00	109.50	30.00	2.1630
OBC-CDC-CC2	70.00	125.00	20.00	2.4420
OSC-CDC-CC2	55.00	109.00	20.00	2.3260
OSC-CDC-OBC	90.00	125.90	160.00	2.2576
OSC-CC2-CC2	75.70	110.10	···	···
OSC-CC2-HA	60.00	109.50	···	···

Table C.4. Bond angle and Urey-Bradley parameters taken from the CHARMM force field. (···) indicate that no Urey-Bradley terms is applied for this atomic pair.

dihedral angle	K_χ/kcal/mole	n	δ
CA-CA-CA-CA	3.100	2	180.00
CA-CA-CA-HP	4.200	2	180.00
HP-CA-HP-CA	2.400	2	180.00
OH1-CT2-CT3-HA	0.160	3	0.00
H-OH1-CT2-CT3	1.300	1	0.00
H-OH1-CT2-HA	0.140	3	0.00
X-CT2-CT3-X	0.160	3	0.00
OB-CD-OS-CT3	0.965	1	180.00
OB-CD-OS-CT3	3.850	2	180.00
X-CD-OS-X	2.050	2	180.00
X-CT3-CD-X	0.000	6	180.00
X-CT3-OS-X	-0.100	3	0.00
CT3-CT2-CT2-CT2	0.150	1	0.00
X-CT2-CT2-X	0.195	3	0.00
O-CC-CT3-HA	0.000	3	180.00
X-CT3-CC-X	0.050	6	180.00
HF1-CF1-CF3-F3	0.158	3	0.00
F3-CF3-CF1-F1	0.185	3	0.00
F3-CF3-CF1-CF3	0.158	3	0.00
CC2-CC2-CC2-CC2	0.150	1	0.00
X-CC2-CC2-X	0.195	3	0.00
OBC-CDC-OSC-CC2	0.965	1	180.00
OBC-CDC-OSC-CC2	3.850	2	180.00
X-CC2-CDC-X	0.000	6	180.00
X-CC2-OSC-X	-0.100	3	0.00
X-CDC-OSC-X	2.050	2	180.00

Table C.5. Dihedral angle parameters taken from the CHARMM27 force field. X can be any atom.

dihedral angle	K_φ/kcal/mole	φ_0
CD-X-X-OB	100.00	0.00
OBC-X-X-CDC	100.00	0.00

Table C.6. Improper dihedral angle parameters taken from the CHARMM27 force field.

atom	ϵ kcal/mole	$R_{min}/2$ A	ϵ^{1-4} kcal/mole	$R_{min}^{1-4}/2$ A
CA	-0.0700	1.9924	···	···
HP	-0.0300	1.3582	···	···
H	-0.0460	0.2245	···	···
OH1	-0.1521	1.7700	···	···
CT2	-0.0550	2.1750	-0.0100	1.9000
CT3	-0.0800	2.0600	-0.0100	1.9000
HA	-0.0220	1.3200	···	···
CD	-0.0700	2.0000	···	···
OB	-0.1200	1.7000	-0.1200	1.4000
OS	-0.1521	1.7700	···	···
CC	-0.0700	2.0000	···	···
O	-0.1200	1.7000	-0.1200	1.4000
CF1	-0.0600	3.8000	···	···
CF3	-0.0390	3.7900	···	···
HF1	-0.0280	2.6400	···	···
F1	-0.1350	3.2600	···	···
F3	-0.0970	3.2440	···	···
CC2	-0.0740	2.1050	···	···
CDC	-0.0700	2.0000	···	···
Lactone parameter set 1				
OBC	-0.3500	1.6450	-0.1200	1.4000
OSC	-0.3000	1.6500	···	···
Lactone parameter set 2				
OBC	-0.6400	1.8800	-0.1200	1.4000
OSC	-0.6400	1.9000	···	···

Table C.7. Lennard-Jones parameters taken from the CHARMM27 force field. ··· indicate that no special parameters are used for 1-4 interactions.

D

Parameters for Reverse Approach Analysis

compound	C	V/mL/mol	T/K
benzene	-55.725	85.524	473.37
ethanol	-23.744	57.834	508.26
methyl acetate	-63.4474	72.731	428.005
n-pentane	-7.48978	135.279	628.365
acetone	-35.0617	72.3087	566.224
γ-butyrolactone	-55.7	85.52	473.0

Table D.1. Pure Component parameters for Huang-O'Connel correlation (Huang and O'Connell, 1987) for reverse approach analysis. The binary mixture parameter k_{12} was st to zero in all the case studies.

T/K	A_{12}	A_{21}	α_{12}	α_{21}	η
benzene(1) - methyl acetate (2)					
303.15	0.260	0.254	...	...	...
ethanol(1) - benzene(2)					
298.15	3.064	1.638	8.420	1.349	2.088
methyl acetate (1) - n-pentane (2)					
298.15	1.5918	1.6359	0.5203	0.6229	0.4517
methyl acetate (1) - acetone (2)					
323.15	0.1079	0.1236	...	...	...
γ-butyrolactone(1) - benzene(2)					
293.15	1.2225	0.3955	0.5450	0.1230	-0.6090
δ-valerolactone(1) - benzene(2)					
293.15	1.5505	0.5168	3.9707	0.8282	...
313.15	1.5419	0.4242	3.1651	0.6758	-1.1225

Table D.2. Modified Margules parameters for reverse approach analysis

compound 1	compound 2	T K	$V_{m,1}$ mL/mol	$V_{m,2}$ mL/mol
benzene	methyl acetate	303.15	80.5217	90.001
benzene	ethanol	298.15	89.5022	58.5159
methyl acetate	n-pentane	298.15	80.000	115.26
methyl acetate	acetone	323.15	82.864	76.673
γ-butyrolactone	benzene	293.15	77.58	91.03

Table D.3. Molar volumes for reverse approach analysis

compound 1	compound 2	T K	A_{12} J/mol	A_{21} J/mol
benzene	methyl acetate	303.15	666.7838	3.1325
benzene	ethanol	298.15	736.0560	6992.556
methyl acetate	n-pentane	298.15	3261.17	1590.14
methyl acetate	acetone	323.15	-397.8058	737.8990
γ-butyrolactone	benzene	293.15	4511.96	-1074.78

Table D.4. Wilson parameters for reverse approach analysis

Compound 1	ethanol
Compound 2	benzene
T/K	298.15
$V_0/mL/mol$	0.09871
$V_1/mL/mol$	0.5741
$V_2/mL/mol$	0.03247
$V_3/mL/mol$	0.3793
$V_4/mL/mol$	0.3591

Table D.5. Handa-Benson V^E correlation parameters for reverse approach analysis. For the rest of the system no V^E term was used in the calculation of the molar volume of the mixtures.

References

A. D. MacKerell, J. and Banavali, N. K. (2000). All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comp. Chem.*, **21**(2), 105Ű–120.

A. D. MacKerell, J.; Bashford, D.; Bellott, M.; R. L. Dunbrack, J.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; W. E. Reiher, I.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wioťrkiewicz-Kuczera, J.; Yin, D. and Karplus, M. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, **102**, 3586–3616.

Abbott, M. M. and van Ness, H. C. (1975). Vapor-Liquid Equilibrium: Part 3. Data Reduction with Precise Expressions for G^E. *AIChE J.*, **21**(1), 62–71.

Aucejo, A.; Montón, J. B.; Muñoz, R. and Wisniak, J. (1996). Double Azeotropy in the Benzene + Hexafluorobenzene System. *Journal of Chemical & Engineering Data*, **41**(1), 21–24.

Aucejo, A.; Loras, S.; Muñoz, R.; Wisniak, J. and Segura, H. (1997). Phase Equilibria and Multiple Azeotropy in the Associating System Methanol + Diethylamine. *Journal of Chemical & Engineering Data*, **42**(6), 1201–1207.

Bayly, C.; Cieplak, P.; Cornell, W. and Kollman, P. (1993). A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *Journal of Physical Chemistry*, **97**(40), 10269–10280.

Ben-Naim, A. (1977). Inversion of the Kirkwood-Buff theory of solutions: Application to the water-ethanol system. *J. Chem. Phys.*, **67**(11), 4884–4890.

Bourasseau, E.; Ungerer, P. and Boutin, A. (2002a). Prediction of Equilibrium Properties of Cyclic Alkanes by Monte Carlo Simulation - New Anisotropic United Atoms Intermolecular Potential - New Transfer Bias Method. *J. Phys. Chem. B*, **106**(21), 5483–5491.

Bourasseau, E.; Ungerer, P.; Boutin, A. and Fuchs, A. H. (2002b). Monte Carlo Simulation of Branched Alkanes and Long-Chain n-alkanes with Anisotropic United Atoms Intermolecular Potential. *Molec. Sim.*, **28**(4), 317–336.

Bourasseau, E.; Haboudou, M.; Boudin, A.; Fuchs, A. H. and Ungerer, P. (2003). New optimization method for intermolecular potentials: Optimization of a new anisotropic united atoms potential for olefins: Prediction of equilibrium properties. *J. Chem. Phys.*, **118**(7), 3020–3034.

Boutard, Y.; Ungerer, P.; Teuler, J.; Ahunbay, M.; Sabater, S.; Perez-Pellitero, J.; Mackie, A. and Bourasseau, E. (2005). Extension of the anisotropic united atoms intermolecular potential to amines, amides and alkanols. *Fluid Phase Equilibria*, **236**(1-2), 25–41.

Briggs, J.; Nguyen, T. and Jorgensen, W. (1991). Monte Carlo simulations of liquid acetic acid and methyl acetate with the OPLS potential functions. *Journal of Physical Chemistry*, **95**(8), 3315–3322.

Burguet, M. C.; Montón, J. B.; Muñoz, R.; Wisniak, J. and Segura, H. (1996). Polyazeotropy in Associating Systems: The 2-Methylpropyl Ethanoate + Ethanoic Acid System. *Journal of Chemical & Engineering Data*, **41**(5), 1191–1195.

Case, F.; Chaka, A.; Friend, D.; Frurip, D.; Golab, J.; Johnson, R.; Moore, J.; Mountain, R.; Olson, J.; Schiller, M. and Storer, J. (2004). The first industrial fluid properties simulation challenge. *Fluid Phase Equilibria*, **217**(1), 1–10.

Case, F.; Chaka, A.; Friend, D.; Frurip, D.; Golab, J.; Gordon, P.; Johnson, R.; Kolar, P.; Moore, J.; Mountain, R.; Olson, J.; Ross, R. and Schiller, M. (2005). The second industrial fluid properties simulation challenge. *Fluid Phase Equilibria*, **236**(1-2), 1–14.

Chen, B. and Siepmann, J. I. (1999). Transferable Potentials for Phase Equilibria. 3. Explicit-Hydrogen Description of Normal Alkanes. *J. Phys. Chem. B*, **103**, 5370–5379.

Chen, B. and Siepmann, J. I. (2000). A Novel Monte Carlo Algorithm for Simulating Strongly Associating Fluids: Applications to Water, Hydrogen Fluoride, and Acetic Acid. *J. Phys. Chem. B*, **104**, 8725–8734.

Chen, B.; Potoff, J. J. and Siepmann, J. I. (2001). Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols. *J. Phys. Chem. B*, **105**, 3093–3104.

Chen, I. J.; Yin, D. and Mackerell, A. (2002). Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. *Journal of Computational Chemistry*, **23**(2), 199–213.

Chialvo, A. A. (1990). Determination of excess Gibbs free energy form computer simulation by the single charging-integral approach. 1. Theory. *J. Chem. Phys.*, **92**(1), 673–679.

Chialvo, A. A. and Haile, J. M. (1987). Determination of Excess Gibbs Free Energy from Computer Simulation: Multiple-Parameter Charging Approach. *Fluid Phase Equilib.*, **37**, 293–303.

Christensen, S. and Olson, J. (1992). Phase equilibria and multiple azeotropy of the acetic acid-isobutyl acetate system. *Fluid Phase Equilibria*, **79**, 187–199.

Christensen, S.; Peters, G. H.; Hansen, F. Y.; O'Connell, J. P. and Abildskov, J. (2007). Generation of Thermodynamic Data for Organic Liquid Mixtures from Molecular Simulations. *Molecular Simulations*, **33**(4-5), 449–457.

Christensen, S.; Peters, G. H.; Hansen, F. Y. and Abildskov, J. (in pressa). Thermodynamic Models from Fluctuation Solution Theory Analysis of Molecular Simulations. *Fluid Phase Equilib., PPEPPD 2007 Special Issue*.

Christensen, S.; Peters, G. H.; Hansen, F. Y.; O'Connell, J. P. and Abildskov, J. (in pressb). Prediction of Bubble Point Pressure Curve for the binary system of ethanol and 1,1,1,2,3,3,3-heptafluoropropane. *Fluid Phase Equilib., IFPSC 2006 Special Issue*.

Cui, S.; Siepmann, J.; Cochran, H. and Cummings, P. (1998). Intermolecular potentials and vapor-liquid phase equilibria of perfluorinated alkanes. *Fluid Phase Equilibria*, **146**(1-2), 51–61.

Dirac, P. A. M. (1958). *Quantum Mechanics*. Oxford University Press, London, fourth edition.

Duncan, W.; Swinton, F.; Sheridan, J. and Gaw, W. (1968). Thermodynamic properties of binary systems containing hexafluorobenzene – 1, 2, 3. *Faraday Society – Transactions*, **64**(543), 637–647.

Dymond, J. and Smith, E. (1980). *The Virial Coefficients of Pure Gases and Mixtures*. Clarendoon Press.

Eckert, F. and Klamt, A. (2001). Validation of the COSMO-RS Method: Six Binary Systems. *Industrial and Engineering Chemistry Research*, **40**(10), 2371.

Eckert, F. and Klamt, A. (2002). Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach. *AIChE Journal*, **48**(2), 369–385.

Errington, J. R. and Panagiotopoulos, A. Z. (1999a). New Intermolecular Potential Model for the n-Alkane Homologous Series. *Journal of Physical Chemistry B - Condensed Phase*, **103**(30), 6314–6322.

Errington, J. R. and Panagiotopoulos, A. Z. (1999b). New intermolecular potential models for benzene and cyclohexane. *J. Chem. Phys.*, **111**(21), 9731–9738.

Foloppe, N. and Mackerell, A. (2000). All-atom empirical force field for nucleic acids. I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, **21**(2), 86–104.

Fredenslund, A.; Jones, R. L. and Prausnitz, J. M. (1975). GROUP-CONTRIBUTION ESTIMATION OF ACTIVITY COEFFICIENTS IN NONIDEAL LIQUID MIXTURES. *AIChE J*, **21**(6), 1086–1099.

Gubbins, K. and Quirke, N., editors (1996). *Molecular Simulations and Industrial Applications*. Gordon and Breach, Amsterdam.

Haile, J. (1986). On the use of computer simulation to determine the excess free energy in fluid mixtures. *Fluid Phase Equilibria*, **26**(2), 103–127.

Haile, J. (1997). *Molecular dynamics simulation : elementary methods*. Wiley.

Handa, Y. and Benson, G. (1979). Volume changes on mixing two liquids: a review of the experimental techniques and the literature data. *Fluid Phase Equilib.*, **3**(2-3), 185–249.

Hayden, J. G. and O'Connell, J. P. (1975). GENERALIZED METHOD FOR PREDICTING SECOND VIRIAL COEFFICIENTS. *Ind Eng Chem Process Des Dev*, **14**(3), 209–216.

Huang, Y. H. and O'Connell, J. P. (1987). CORRESPONDING STATES CORRELATION FOR THE VOLUMETRIC PROPERTIES OF COMPRESSED LIQUIDS AND LIQUID MIXTURES. *Fluid Phase Equilib.*, **37**, 75–84.

Hwang, S.-C. and Robinson, R. L. J. (1977). VAPOR-LIQUID EQUILIBRIA AT 25 degree C FOR NINE ALCOHOL-HYDROCARBON BINARY SYSTEMS. *J Chem Eng Data*, **22**(3), 319–325.

Ihmels, E. C. and Gmehling, J. (2002). Liquid Densities of γ-Butyrolactione and N-methyl-2-pyrolidone from 273 to 473 K and at Pressures up to 40 MPa. *J. Chem. Eng. Data*, **47**, 1547–1552.

Jorgensen, W. and Swenson, C. (1985). Optimized intermolecular potential functions for amides and peptides. Structure and properties of liquid amides. *Journal of the American Chemical Society*, **107**(3), 569–578.

Jorgensen, W. L. (1981a). Transferable Intermolecular Potential Functions for Water, Alcohols, and Ether. Application to Liquid Water. *J. Am. Chem. Soc.*, **103**, 335–340.

Jorgensen, W. L. (1981b). Transferable Intermolecular Potential Functions. Application to Liquid Methanol Including Internal Rotation. *J. Am. Chem. Soc.*, **103**, 341–345.

Jorgensen, W. L. (1981c). Simulation of Liquid Ethanol Including Internal Rotation. *J. Am. Chem. Soc.*, **103**, 345–350.

Jorgensen, W. L. (1986). Optimized Intermolecular Potential Functions for Liquid Alcohols. *J. Phys. Chem.*, **90**, 1276–1284.

Jorgensen, W. L.; Madura, J. D. and Swenson, C. J. (1984). OPTIMIZED INTERMOLECULAR POTENTIAL FUNCTIONS FOR LIQUID HYDROCARBONS. *J. Am. Chem. Soc.*, **106**(22), 6638–6646.

Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K. and Schulten, K. (1999). NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comp. Phys.*, **151**(1), 283Ű–312.

Kamath, G.; Robinson, J. and Potoff, J. J. (2006). Application of TraPPE-UA force field for determination of vapor-liquid equilibria of carboxylate esters. *Fluid Phase Equilibria*, **240**(1), 46–55.

Kao, C.-P. C.; Paulaitis, M. and Yokozeki, A. (1997). Double azeotropy in binary mixtures of NH_3 and CHF_2CF_3. *Fluid Phase Equilibria*, **127**(1), 191–203.

Kirkwood, J. G. and Buff, F. P. (1951). The statistical Mechanical Theory of Solutions. I. *J. Chem. Phys.*, **19**(6), 774–777.

Klamt, A. (1995). Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *Journal of Physical Chemistry*, **99**(7), 2224–2235.

Klamt, A. (2005). *COSMO-RS from quantum chemistry to fluid phase thermodynamics and drug design*. Elsevier.

Klamt, A. and Eckert, F. (2004). Prediction of vapor liquid equilibria using COSMOtherm. *Fluid Phase Equilibria*, **217**(1), 53–57.

Klamt, A.; Eckert, F.; Hornig, M.; Beck, M. E. and Burger, T. (2002). Prediction of Aqueous Solubility of Drugs and Pesticides with COSMO-RS. *Journal of Computational Chemistry*, **23**(2), 275–281.

Klaus, R. and Van Ness, H. (1967). Orthogonal polynomial representation of thermodynamic excess functions. *Chemical Engineering Progress Symposium Series*, **63**(81), 88–104.

Klein, A. and Svejda, P. (1995). Isothermal Vapor-Liquid Equilibria and Excess Volumes of Binary Mixtures of Benzene + γ- Butyrolactone, γ-Valerolactone, δ-Valerolactone, or ϵ-Caprolactone at 293.15 K and 313.15 K. *ELDATA Int. Electron. J. Phys. Chem. Data*, **1**, 87–94.

Kranias, S.; Pattou, D.; Levy, B. and Boutin, A. (2003). An optimized potential for phase equilibria calculation for ketone and aldehyde molecular fluids. *Physical Chemistry Chemical Physics*, **5**(19), 4175–4179.

Laso, M.; de Pablo, J. J. and Suter, U. W. (1992). Simulation of phase equilibria for chain molecules. *J. Chem. Phys.*, **97**(4), 2817–2819.

Leu, A. D. and Robinson, D. B. (1991). Vapor-liquid equilibrium in selected binary systems of interest to the chemical industry. In *Experimental results for phase equilibria and pure components properties*, number 1 in AIChE data series, pages 1–5.

Levy, B. and Enescu, M. (1998). Theoretical study of methylene blue: a new method to determine partial atomic charges; investigation of the interaction with guanine. *THEOCHEM*, **432**(3), 235–245.

Liu, Z.; Xu, Y.; Saladino, A. C.; Wymore, T. and Tang, P. (2004). Parametrization of 2-Bromo-2-Chloro-1,1,1-Trifluoroethane (Halothane) and Hexafluoroethane for Nonbonded Interactions. *J. Phys. Chem. A*, **102**, 781–786.

Lopes, J. N. C. and Tildesley, D. J. (1997). Multiphase equilibria using the Gibbs ensemble Monte Carlo method. *Molec. Phys.*, **92**(2), 187–195.

Lu, B. C.-Y.; Ishikawa, T. and Benson, G. C. (1990). Isothermal Vapor-Liquid Equilibria for n-Hexane-Methyl Methacrylate, Methyl n-Propyl Ketone-Acetic Acid, n-Pentane-Methyl Acetate, and Ethyl Acetate-Acetic Anhydride. *J. Chem. Eng. Data*, **35**, 331–334.

Martin, M. and Biddy, M. (2005). Monte Carlo molecular simulation predictions for the heat of vaporization of acetone and butyramide. *Fluid Phase Equilibria*, **236**(1-2), 53–57.

Martin, M. and Frischknecht, A. (2006). Using arbitrary trial distributions to improve intramolecular sampling in configurational-bias Monte Carlo. *Molecular Physics*, **104**(15), 2439–2456.

Martin, M. and Thompson, A. (2004). Industrial property prediction using Towhee and LAMMPS. *Fluid Phase Equilibria*, **217**(1), 105–110.

Martin, M. G. and Siepmann, J. I. (1998). Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes. *J. Phys. Chem. B*, **102**(14), 2569–2577.

Martin, M. G. and Siepmann, J. I. (1999). Novel Configurational-Bias Monte Carlo Method for Branched Molecules. Transferable Potentials for Phase Equilibria. 2. United-Atom Description of Branched Alkanes. *J. Phys. Chem. B*, **103**, 4508–4517.

Matteoli, E. and Lepori, L. (1984). Solute-solute interactions in water. II. An analysis through the Kirkwood-Buff integrals for 14 organic solutes. *J. Chem. Phys.*, **80**(6), 2856–2863.

Matteoli, E. and Mansoori, G. (1995). A simple expression for radial distribution functions of pure fluids and mixtures. *J. Chem. Phys.*, **103**(11), 4672–4677.

Matteoli, E. and Mansoori, G. A. (1990). *Fluctuation Theory of Mixtures, Advances in Thermodynamics*, volume 2. Taylor and Francis, New York.

McQuarrie, D. A. (2000). *Statistical mechanics*. University Science Books.

Nagai, Y. and Isii, N. (1935). Studies on volatility of fuels containing ethyl alcohol. *Society of Chemical Industry, Japan – Journal*, **38**(1).

Nagata, I.; Ohta, T.; Takahashi, T. and Gotoh, K. (1973). Thermodynamic properties of methyl acetate-benzene and methyl acetate-cyclohexane mixtures. *J. Chem. Eng. Japan*, **6**, 129–134.

Nath, S. K. and de Pablo, J. J. (2000). Simulation of vapour-liquid equilibria for branched alkanes. *Molec. Phys.*, **98**(4), 231–238.

O'Connell, J. (1971a). Thermodynamic properties of solutions based on correlation functions. *Molec. Phys.*, **20**(1), 27–33.

O'Connell, J. P. (1971b). Molecular Thermodynamics of Gases in Mixed Solvents. *AIChE J.*, **17**(3), 658–663.

Olson, J. (1981). Ebuliiometric Determination of PTx Data and G^E for Acetone + Methyl Acetate from 20 to 60 °C. *J. Chem. Eng. Data*, **26**, 58–64.

Panagiotopoulos, A. Z. (1987). Direct determination of phase coexistience properties of fluids by Monte Carlo simulation in a new ensemble. *Molec. Phys.*, **61**(4), 813–826.

Panagiotopoulos, A. Z. (1992). Direct determination of fluid phase equilibria by simulation in the Gibbs ensemble: a review. *Molec. Sim.*, **9**, 1–23.

Panagiotopoulos, A. Z.; Quirke, N.; Stapleton, M. and Tildesley, D. J. (1988). Phase equilibria by simulation in the Gibbs ensemble: Alternative derivation, generalization and application to mixture and membrane equilibria. *Molec. Phys.*, **63**(4), 527–545.

Potoff, J.; Errington, J. and Panagiotopoulos, A. (1999). Molecular simulation of phase equilibria for mixtures of polar and non-polar components. *Molecular Physics*, **97**(10), 1073–1083.

Putnam, R.; Taylor, R.; Klamt, A.; Eckert, F. and Schiller, M. (2003). Prediction of infinite dilution activity coefficients using COSMO-RS. *Industrial and Engineering Chemistry Research*, **42**(15), 3635–3641.

Salvi-Narkhede, M.; Wang, B.-H.; Adcock, J. and van Hook, W. A. (1992). Vapor pressures, liquid molar volumes, vapor non-ideality, and critical properties of some partially fluorinated ethers ($CF_3OCF_2CF_2H$, CF_3OCF_2H, and CF_3OCH_3), some perfluoroethers ($CF_3OCF_2OCF_3$, c-$CF_2OCF_2OCF_2$, and c-$CF_2CF_2CF_2O$), and of CHF_2Br and CF_3CFHCF_3. *J. Chem. Thermodynamics*, **24**, 1065–1075.

Sapei, E.; Zaytseva, A.; Uusi-Kyyny, P.; Keskinen, K. I. and Aittamaa, J. (submitted). Vapor-Liquid Equilibrium for Binary System of Thiophene + 2,2,4- Trimethylpentane at 353.15 and 343.15 K and Thiophene + 2-Ethoxy-2- Methylpropane at 343.15 and 333.15 K. *Fluid Phase Equilib., PPEPPD 2007 Special Issue.*

Slusher, J. T. (1998). Estimation of infinite dilution activity coefficient in aqueous mixtures via molecular simulation. *Fluid Phase Equilib.*, **153**, 45–61.

Slusher, J. T. (1999). Infinite dilution activity coefficient in hydrogen-bonded mixtures via molecular dynamics: the water/methanol system. *Fluid Phase Equilib.*, **154**, 181–192.

Smit, B. and Frenkel, D. (1989). Calculation of the chemical potential on the Gibbs ensemble. *Molec. Phys.*, **68**(4), 951–958.

Smith, G. and Jaffe, R. (1996). Quantum chemistry study of conformational energies and rotational energy barriers in n-alkanes. *Journal of Physical Chemistry*, **100**(48), 18718–18724.

Smith, J.; van Ness, H. and Abbott, M. (2005). *An Introduction to Chemical Engineering Thermodynamics*. McGraw-Hill, 7th edition.

Stubbs, J. M.; Chen, B.; Potoff, J. J. and Siepmann, J. I. (2001). Monte Carlo calculations for the phase equilibria of alkanes, alcohols, water, and their mixtures. *Fluid Phase Equilib.*, **183–184**, 301–309.

Stubbs, J. M.; Potoff, J. J. and Siepmann, J. I. (2004). Transferable Potentials for Phase Equilibria. 6. United-Atom Dscription for Ethers, Glycols, Ketones, and Aldehydes. *J. Phys. Chem. B*, **108**, 17596–17605.

Toxvaerd, S. (1990). Molecular dynamics calculation of equation of state of alkanes. *J. Chem. Phys.*, **93**(6), 4290–4295.

Toxvaerd, S. (1997). Equation of state of alkanes II. *J. Chem. Phys.*, **107**(13), 5197–5204.

Ungerer, P.; Beauvais, C.; Delhomelle, J.; Boutin, A. and Rousseau, B. (2000). Optimization of the anisotropic united atoms intermolecular potential for n-alkanes. *J. Chem. Phys.*, **112**(12), 5499–5510.

Verdier, P. H. and Stockmayer, W. H. (1962). Monte Carlo Calculations on the Dynamics of Polymers in Dilute Solution. *J. Chem. Phys.*, **36**, 227.

Vlugt, T. J. H.; Martin, M. G.; Smit, B.; Siepmann, J. I. and Krishna, R. (1992). Improving the efficiency of the configurational-bias Monte Carlo algorithm. *Molec. Phys.*, **94**(4), 727–733.

Wang, S.; Lin, S.-T.; Chang, J.; Goddard, W. A. I. and Sandler, S. I. (2006). Reply to the Comment on "Application of the COSMO-SAC-BP Solvation Model to Predictions of Normal Boiling Temperatures for Environmentally Significant Substances". *Industrial and Engineering Chemistry Research*, **45**(10), 3767.

Weerasinghe, S. and Smith, P. E. (2003). Kirkwood-buff derived force field for mixtures of acetone and water. *J. Chem. Phys.*, **118**(23), 10663–10670.

Wick, C. D.; Martin, M. G. and Siepmann, J. I. (2000). Transferable Potentials for Phase Equilibria. 4. United-Atom Description of Linear and Branched Alkenes and Alkylbenzenes. *J. Phys. Chem. B*, **104**(33), 8008–8016.

Wick, C. D.; Siepmann, J. I. and Schure, M. R. (2003). Temperature Dependence of Transfer Properties: Importance of Heat Capacity Effects. *J. Phys. Chem. B*, **107**, 10623–10627.

Wick, C. D.; Stubbs, J. M.; Rai, N. and Siepmann, J. I. (2005). Transferable Potentials for Phase Equilibria. 7. Primary, Secondary and Tertiary Amines, Nitroalkanes and Nitrobenzene, Nitriles, Amides, Pyridine, and Pyrimidine. *J. Phys. Chem. B*, **109**, 18974–18982.

Wooley, R. J. and O'Connell, J. P. (1991). A database of fluctuation thermodynamic properties and molecular correlation function integrals for a variety of binary liquids. *Fluid Phase Equilib.*, **66**, 233–261.

Zhang, L. and Ilja Siepmann, J. (2005). Pressure dependence of the vapor-liquid-liquid phase behavior in ternary mixtures consisting of n-alkanes, n-perfluoroalkanes, and carbon dioxide. *Journal of Physical Chemistry B*, **109**(7), 2911–2919.

List of Abbreviations

AA All Atom

AUA Anisotropic United Atom

GCE Grand Canonical Ensemble

GEMC Gibbs Ensemble Monte Carlo

GFEI Gibbs Free Energy Integration

IFPSC Industrial Fluid Phase Simulation Challenge

LLE liquid-liquid equilibrium

MD Molecular Dynamic

MM Molecular Modeling

QM Quantum Mechanic

SCD Screening Charge Density

SLE solid-liquid equilibrium

TCFI Total Correlation Function Integral

TraPPE-EH Transferable Potentials for Phase Equilibria - Explicit Hydrogen

TraPPE-UA Transferable Potentials for Phase Equilibria - United Atom

VLE vapor-liquid equilibrium